DELAYED VERSUS IMMEDIATE CORRECTIVE FEEDBACK ON ORALLY PRODUCED PASSIVE ERRORS IN ENGLISH

by

Paul Quinn

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education
University of Toronto

DELAYED VERSUS IMMEDIATE CORRECTIVE FEEDBACK ON ORALLY PRODUCED PASSIVE ERRORS IN ENGLISH

Doctor of Philosophy, 2014
Paul Quinn
Graduate Department of Curriculum, Teaching and Learning
University of Toronto

Abstract

Research demonstrating the beneficial effects of corrective feedback (CF) for second language (L2) learning (e,g., Li, 2010) has almost invariably resulted from studies in which CF was provided immediately. Yet teachers are often encouraged to delay CF to avoid interrupting learners (Harmer, 2001). This study investigates how differences in the timing of CF on oral production affect L2 learning and learners' reactions to CF.

Theoretically, Immediate CF may facilitate L2 development by allowing learners to immediately compare their errors to accurate models (i.e., recasting, e.g., Doughty, 2001). The effectiveness of Immediate CF has also been linked to skill acquisition theory because some CF (i.e., prompting) is hypothesized to help learners proceduralize their L2 knowledge (Ranta & Lyster, 2007). This thesis introduces additional theoretical explanations to explain the effectiveness of both Immediate and Delayed CF. For example, reactivation and reconsolidation theory (Nader & Einarsson, 2010) holds that long-term mental representations are susceptible to change when they are recalled. Thus, both Immediate and Delayed CF may help learners alter their incorrect mental representations of language features if that CF reminds learners of those incorrect representations and provides them with accurate models.

In a laboratory-based study, 90 intermediate-level adult ESL learners were randomly assigned to Immediate, Delayed, and No CF conditions. Learners took three

pre-tests to measure their knowledge of the English passive construction: an aural grammaticality judgment test (AGJT), an oral production test (OPT), and a written error correction test (ECT). Next, they received some brief instruction on the passive. Learners then completed three communicative tasks in which the CF conditions were provided. These tasks were followed by immediate and delayed post-tests. Learners' reactions to CF were elicited with a questionnaire.

Mixed-design one-way ANOVAs revealed statistically significant improvement for all conditions over time on all measures, but no statistically significant differences between conditions. The questionnaires revealed that learners prefer Immediate CF, but that Immediate CF may constrain CF noticeability and learners' independence, while Delayed CF may cause anxiety or embarrassment. In summary, altering the timing of CF did not differentially affect L2 development, but it did elicit different reactions from learners.

Acknowledgements

First and foremost, I am exceedingly grateful to my incredible supervisor and mentor, Dr. Nina Spada, for her constant wise counsel when decisions had to be made, her extraordinary dedication to all aspects of my dissertation, and her indefatigable support of me and my work. Also, I want to thank my committee members, Dr. Alister Cumming and Dr. Julie Kerekes; my internal-external examiner, Dr. Eunice Jang; and my external examiner, Dr. Roy Lyster. It will be my continuing aspiration to make all of you proud in my future endeavours.

I also wish to thank many others for their assistance, support, and guidance: Dr. Mohammed Al-Alawi, Dr. Rachael Cayley, Greta Camase, the Comps Study Group, Dr. Olesya Falenchuk, Dr. Jane Freeman, Dr. Nobuhiro Kamiya, Robert Kohls, Dr. Shawn Loewen, Roy McCool, study participants, Dan Riccardi, Dr. Kazuya Saito, the Spada Research Group (past and present members), Anthony Stille, Dr. Saskia Stille, Dr. Wataru Suzuki, Dr. Merrill Swain, and Dr. Choongil Yoon.

I am very grateful as well for the funding that I received from the following organizations that helped me complete my doctoral studies: The Government of Ontario, for the Ontario Graduate Scholarship; Language Learning Journal, for the Dissertation Grant; The Association of Japanese Canadian Businesses and Professionals, for the Shinki-Kai Scholarship; and University of Toronto/OISE, for the Graduate Funding Grants.

I also would like to thank my family for forgiving my absences at family events when deadlines loomed. Furthermore, I am particularly grateful to the Nekonomori clan,

especially Maui, who has provided unwavering support and companionship and uncomplainingly endured innumerable drafts read aloud in his presence.

Finally, I thank my wife, the love of my life and the most important person in the world to me, for her inestimable sacrifices and invaluable support during my graduate studies. Kinuyo, *we* did it!

Table of Contents

List of Tables

List of Appendices

Dedication

To my wife Kinuyo

Chapter One

Introduction

Qui tacet consentire videtur; ubi loqui debuit ac potuit (One who is silent, when one should and might have spoken, is seen to consent). When language teachers are silent about learners' inaccurate language use, learners naturally assume that teachers' silence indicates that they have made no errors. There is no negative evidence to lead learners to believe otherwise. In second language (L2) education, teachers' knowledge of the L2, their teaching experience, and their positions of power in the classroom make them the ones who are expected to present that negative evidence, to correct errors, to speak when they should. However, when should they speak? Is there an optimal time to correct errors? What results from correcting errors immediately rather than after some delay? These questions –of relevance to L2 learners, teachers, and researchers– inspired this doctoral dissertation's investigation of the effects on second language acquisition (SLA) of altering the timing of corrective feedback (CF).

Following decades of CF research, there is now a general consensus in the SLA literature that oral CF is beneficial to L2 learning (e.g., Li, 2010; Lyster & Saito, 2010; Lyster, Saito & Sato, 2013; Russell & Spada, 2006). When L2 learners make grammar errors in a spoken communicative task, L2 teachers can provide CF immediately during the task or delay CF until some time after the task has ended. CF on spoken grammar errors has thus far been found to be most effective when learners are aware of its corrective intent as it draws their attention to form within communicative practice. However, to avoid interrupting spoken tasks, teacher-training texts often encourage teachers to provide CF after, rather than during, communicative practice (Bartram & Walt,

1991; Edge, 1989; Harmer, 2001; Hedge, 2000). In fact, little is known about whether such delayed CF results in the same development of L2 grammatical knowledge that results from immediately-provided CF.

In the decade that I taught English and trained English teachers in Japan, I provided both immediate and delayed CF, and trained teachers to do the same. Communication was the main focus of our lessons, so we limited immediate CF to what we called "echo-correcting," what is known in SLA research as recasting (e.g., Goo & Mackey, 2013; Lyster & Ranta, 1997; Nicholas, Lightbown & Spada, 2001). I was trained to instruct teachers that the primary value of this type of correction through paraphrase was that it allowed them to provide CF without breaking the flow of learners' communication. I was instructed to teach them that the delayed CF provided after communicative practice was primarily responsible for facilitating L2 development and that the immediate CF through recasting might contribute, but that its main goal was to be, as Ellis and Sheen (2006) phrase it, "a conversational lubricant" (p. 585).

However, when I began studying the SLA research literature in 2005, I was stunned by how rarely delayed CF (of oral, not written, production) was discussed at a time when CF was such a popular topic of investigation. Discussion of the timing of CF had appeared frequently in the literature of the 1970's (e.g., Allwright, 1975; Chastain, 1971; Chaudron, 1977a; Cohen, 1975; Fanselow, 1977; Hendrickson, 1978; Long, 1977), but since that time and into the early 2000's, few researchers (e.g., Doughty, 2001; Loewen, 2004) even broached the subject. In fact, during that period, no published empirical research on the effect of delayed CF was conducted, to my knowledge. Instead, attention was focused on types of immediately-delivered CF rather than on the timing of

CF (e.g., Ellis, Basturkmen, & Loewen, 2001; Lyster & Ranta, 1997; Sheen, 2004), with recasts being by far the most popular type of CF under investigation (e.g., Doughty & Varela, 1998; Mackey, Gass, & McDonough, 2000; Mackey & Philp, 1998; Nicholas et al., 2001; Ohta, 2000; Oliver, 1995). This pattern of CF research, with its near exclusive focus on CF type, has changed little since that time (e.g., Ammar & Spada, 2006; Ellis, Loewen & Erlam, 2006; Lyster & Izquierdo, 2009; Yang & Lyster, 2010). The "heated debate over which feedback type is more effective" (p. 348) as Li (2010) refers to it, has only intensified in the first two decades of the 21st century (e.g., Goo & Mackey, 2013; Long, 2007; Lyster, 2004; Lyster & Ranta, 2013).

A detrimental effect of this nearly singular focus on CF type is that, in contemporary CF research, it seems apriori that CF is to be provided immediately. It is as though, as Hunter (2007) has argued, "an assumption has been made that feedback can only be effective if offered 'online', that is, in the immediate context of the error" (p. 49). However, the positive results from the very few investigations of the developmental effects of delayed CF (Hunter, 2011; Sheen, 2012; Siyyari, 2005; Varnosfadrani, 2006) offer some counter-evidence to that assumption. Nonetheless, as Roediger (2013) notes about entrenched ideas in educational psychology, "once an idea has taken hold, it is hard to root out....[even] an idea supported by, at best, only indirect evidence, and at worst no rigorous experimental evidence" (p. 2). Therefore, more research is required to investigate whether there is any difference in the L2 development that results from immediate and delayed CF. Knowledge about the L2 development that results from CF will remain vastly incomplete if researchers continue to focus solely upon *how* CF should be provided and fail to adequately explore *when* it should be provided. Moreover, if SLA

research is to serve not only researchers but also teachers and learners, as I contend it should, then surely researchers must investigate whether there is any empirical evidence that can be found to justify the largely unsupported prescriptions about the timing of CF in the pedagogical literature

Given this state of affairs, it is not surprising that calls are being made in the SLA literature to compare the developmental effects of immediate and delayed CF (e.g., Ellis, 2012; Lyster et al., 2013). This dissertation responds to those calls with an experimental study guided by the following three research questions:

1) Is there a difference in the L2 grammatical development that results from instruction that includes CF and instruction that does not?

2) Is there a difference in the L2 grammatical development that results from instruction that includes Immediate CF[1] and instruction that includes Delayed CF?

3) How do learners react to instruction that includes Immediate CF or Delayed CF?

These research questions are addressed through a laboratory-based study of 90 participants who were randomly assigned to Immediate, Delayed and No CF conditions. The study utilized a pre- to immediate post- to delayed post-test experimental research design that featured an instructional intervention in which the CF conditions were provided between the pre- and immediate post-tests. A post-instruction questionnaire was employed to investigate participants' reactions to the CF in the instructional treatment. Finally, statistical and thematic analyses of the data were conducted to determine the responses to the research questions.

[1] Capitalization will be used for the three treatment conditions throughout the thesis.

This thesis describes the study and discusses its results. The first chapter has served to explain the rationale behind this investigation and introduce the research questions guiding it. In the second chapter, I argue that the subject of timing in CF has been marginalized in CF research literature, but that the evidence that CF facilitates L2 development implies that development should result not only from immediate CF but also delayed CF. Then, I review and expand upon the theoretical literature that underpins CF to suggest theoretical explanations for why both immediate and delayed CF should facilitate L2 development. Finally, I review the few studies that have focused primarily upon the timing of CF and explain why they indicate that more research into the timing of CF is required. In the third chapter, I explain the method used to investigate the research questions, including the instructional treatment, language measures, and post-instruction questionnaire. In the fourth chapter, I describe how the data from the language measures and questionnaire were analyzed and provide the results. Finally, in the fifth chapter, I discuss how the results respond to the research questions, and I provide potential explanations for those outcomes. I also discuss the pedagogical and theoretical implications of the study and provide a final summary of the thesis.

Chapter Two

Empirical and Theoretical CF Research

In this chapter, I review the empirical and theoretical CF literature to demonstrate the need for more research into the developmental effects related to the timing of CF. First, I discuss how the issue of timing of CF has been marginalized. Next, I review the CF research about types of immediate CF that facilitate L2 development. I argue that the empirical findings from that research imply that delayed CF should also result in L2 development. Then, I review and expand upon the current theoretical explanations for the effectiveness of CF to provide theoretical reasons for why L2 development should result from both immediate and delayed CF. Finally, I demonstrate the need for more research into the timing of CF by explaining the limitations of the few studies that have focused primarily on the subject.

2.1. The Timing of Corrective Feedback: A Marginalized Issue

In the instructed SLA research literature, the issue of the timing of CF has long been marginalized, and interest in it has only recently re-emerged. Prior to the era of communicative language teaching (CLT), the timing of CF was regularly discussed, with some theorists suggesting that CF be provided immediately and others that it be delayed. However, since that time, the discussion of CF timing has greatly diminished. Research studies that focus on the timing of CF are extremely rare in comparison to the large number of studies that focus on type of CF (i.e., research concerned with how the effectiveness of CF varies depending on the degree of explicitness or implicitness or function of the types of CF that are provided).

Beliefs about the timing of CF have been greatly influenced by changes in thinking about the way languages are learned. In the 1950s and 60s, theorists urged that error be avoided like sin, and that if it had to be dealt with at all that it be dealt with immediately (Hendrickson, 1978). Under the influence of behaviourism, theorists like Brooks (1960) urged that as little time as possible be allowed between the time learners committed errors and the time teachers provided a correct model. Theorists feared the longer learners were exposed to an unaddressed incorrect linguistic behaviour, the more likely it was that the behaviour would become ingrained.

A shift in thinking took hold with Corder's (1967) seminal recognition of the importance of errors as illustrations of the development learners were making with the L2. Teachers were then urged to analyze errors and determine what hypotheses learners might be testing about the L2 through those errors. Fanselow (1977) suggested that this analysis might require novice teachers to delay CF up to a day to analyze what an error was indicating and how to address it appropriately. Along the same lines, Chastain (1971) suggested that teachers might review common mistakes after communicative activities had ended. Allwright (1975), on the other hand, argued that teachers must not wait but must analyze and provide instant analysis and feedback.

Long (1977) observed that arguments exist that both oppose and support delaying CF. He cited psychological research (Mackworth, 1950) which indicated that the value of feedback decreased as the amount of time between the commission of the error and the corrective response increased, and that when researchers provided feedback after each incorrect response (or each failure to respond) participants remained more vigilant about their accuracy than was normally the case.

However, Long (1977) also recognized that delayed correction has the advantage of avoiding the inhibiting effects of interrupting learners. In explaining what he meant by "delayed", Long drew upon Chaudron's (1977b) unpublished research on how error was treated in French immersion classes. Long noted that according to Chaudron (1977b), correction can be either "delayed" or "postponed." Long explains that "delayed correction" referred to correction provided "after the (apparent) completion of the student's utterance" and "postponed CF" referred to correction "provided at some future time including, for example, in the course of a future lesson" (p. 290). Holley and King (1971) found that delayed correction, as defined above, allowed students to self-correct and benefit more from the correction that teachers offered. However, the notion of delayed correction with a delay of merely 5 to 10 seconds (Long, 1977) seems to have been rejected by subsequent CF researchers. Contemporary CF research has conflated 5-10 second-delayed CF with immediate CF.[2] In this dissertation, unless stated other wise, "delayed CF" refers to CF provided after a communicative activity, as is standard in contemporary CF research (e.g., Hunter, 2012; Rolin-Ianziti, 2006, 2010; Siyyari, 2005; Varnosfadrani, 2006). Thus, Delayed CF in this dissertation is more in line with what Long (1977) and Chaudron (1977b) referred to as "postponed correction," and with the type of correction that Fanselow (1977) and Chastain (1971) recommended.

Discussion of the timing of CF greatly diminished with the onset of the era of CLT. In CLT, the ability to communicate one's intention is emphasized, while the focus on doing so accurately receives less attention. There are multiple explanations for the genesis and subsequent popularity of CLT (Spada, 2006), but arguably the prime

[2] To my knowledge, no contemporary researchers draw a distinction between CF that interrupts learners in mid-speech and CF that allows them to complete their utterance. Both are simply referred to as CF.

motivator was the widespread disenchantment with teaching languages through the explanation and memorization of grammar rules, the style of teaching that Long (1991) refers to as a focus on forms. With the onset of CLT, the question "When should CF be provided?" was displaced by the questions "Should CF be provided?", and "If CF should be provided, how should that be done without interfering with the communication of meaning?"

One response to the question "Should CF be provided?" came from Krashen (1985), who argued in his influential input hypothesis that CF was not only unnecessary but also detrimental to the acquisition process[3]. Krashen was inspired by the apparent lack of need for explicit correction of language errors in successful first language (L1) acquisition. Krashen (1982) argued that learners have negative emotional reactions to being corrected. He hypothesized that correction raises the 'affective filter' which, he claimed, blocks the unconscious acquisition of an L2. For Krashen, all that is necessary for L2 acquisition to occur is "comprehensible input," that is, input just beyond an L2 learner's ability to understand. In time, logical arguments and empirical evidence convinced most SLA researchers that the sole focus on meaning prescribed in the input hypothesis was not sufficient for near-native like L2 acquisition to occur; some attention to language form was necessary (White, 1987). Empirically, the findings from research into immersion programs indicated that attention to meaning alone was not sufficient for attaining native-like proficiency in an L2, especially in terms of the acquisition of L2 grammar (Swain, 1985).

[3] Krashen appears to believe that learners can use CF to learn *about* the nature of a linguistic system but not to acquire it. In Krashen's input hypothesis (1985), the acquisition of a language is completely distinct from learning the rules of a language.

The question "Should CF be provided?" lost intellectual ground to the question "How should CF be provided?" Research influenced by the interaction hypothesis (Long, 1983) was one influential way of addressing that question. Interactionists hypothesize that L2 learners' speech is prone to errors, and that the optimal time for SLA is when those errors cause breakdowns in communication to occur between L2 learners and interlocutors. Such communication breakdowns force L2 learners and their interlocutors to negotiate what failed to be comprehended until the meaning is understood. These negotiations of meaning employ particular types of conversation devices such as clarifications requests and comprehension checks. Researchers initially focused on how interaction was affected by the types of interlocutors (e.g., Ellis, 1985; Gass & Varonis, 1985), types of tasks (e.g., Doughty & Pica, 1986), and how interaction led to learners' modifying their output (Pica, 1988; Pica, Holliday, Lewis, & Morgenthaler, 1989). Eventually, however, the main focus of interaction research was how effective certain types of interaction (e.g., recasts) were in leading to L2 development. In Long's (1996) revised interaction hypothesis, interaction is viewed as a source of negative evidence that acts as CF which can facilitate L2 development. Importantly, Lyster (1994) argued that negotiation must not be limited to meaning because teachers also negotiate language form, by using negotiation moves as CF to push learners to reformulate inaccurate utterances even when the students' intended meaning is clear. The types of CF moves used in such error-instigated negotiations (of meaning or form) are still the primary variables in CF research today. However, these types of CF are now more commonly referred to as kinds of form-focused instruction (FFI) (Spada, 1997) – a meaning-focused style of instruction in which attention is occasionally directed to language form. Because these types of CF

have always been provided immediately in FFI research, there was little motivation to discuss the timing of CF in any significant manner until the new millennium began.

The timing issue was re-introduced by Doughty (2001) who argued that in order for CF to be effective it needed to be provided immediately within the context of meaning-based interaction. This argument is explained in more detail below in Section 2.3.1.2. Other work of a more empirical nature followed, including a few observational studies of the timing of CF in second/foreign language classroom settings. One study indicated that classroom teachers provide both immediate and delayed CF (Loewen, 2004), while three others described how classroom teachers provided delayed CF (Hunter, 2012; Rolin-Ianziti, 2006, 2010). The finding that teachers in real classrooms delay CF is not surprising because as noted above, teacher-training texts often recommend that teachers delay CF to avoid interrupting communicative activities (Bartram & Walt, 1991; Edge, 1989; Harmer, 2001; Hedge, 2000).

Although attention to the timing of CF has been limited in the contemporary SLA literature, the findings from the research into types of CF imply that delayed CF, like immediate CF, should facilitate L2 development. Furthermore, findings from the few studies that have investigated the developmental effects of the timing of CF (Hunter, 2011; Sheen, 2012; Siyarri, 2005; Varnosfadrani, 2006) are more supportive of that conclusion than they are of the implication from Doughty (2001) that L2 development from CF is restricted to CF that is provided immediately. The findings from those studies serve as evidence that delayed CF is effective and perhaps even as effective as immediately-provided CF. However, as will be discussed in Section 2.4., the limitations

of those studies indicate that more research is needed to learn about the L2 development that results from altering the timing of CF.

2.2. The Effectiveness of Immediate CF Implies Delayed CF Should be Effective

FFI researchers who have investigated the effects of various types of CF on the development of L2 grammatical knowledge have almost without exception investigated CF that has been provided during (not after) communicative practice. These researchers have found that 1) CF facilitates the development of L2 grammatical knowledge (Li, 2010; Lyster & Saito, 2010; Lyster et al., 2013) and 2) learners' awareness of the corrective intent of CF is important in that facilitation (Nicholas et al., 2001; Sheen, 2010a). Because CF that is provided during communicative practice facilitates the development of L2 grammatical knowledge, it seems quite plausible that delayed CF would also do so. Furthermore, it seems quite probable that delayed CF would facilitate the development of L2 grammatical knowledge because the corrective intent of CF that is provided after a communicative activity is unambiguous to learners. Below, I briefly review the evidence that demonstrates that CF provided within communicative practice facilitates the development of L2 grammatical knowledge. After doing so, I review the evidence that learners' awareness of the corrective intent of CF is important in that facilitation.

Prior to the widespread use of CLT, error treatment was one of only two universal characteristics found in a comparison of language teaching methods (Krashen & Seliger, 1975)[4]. However, as the influence of CLT grew, teachers' practices divided: some continued to practice some form of CF while others believed that learning language through communication eliminated the need for error treatment (Nicholas et al., 2001).

[4] The other characteristic was discrete point presentation.

As explained above, some researchers discouraged teachers from using CF in the L2 classroom because they claimed that CF may make students defensive, undermine their confidence, and give them a negative attitude toward language learning (Krashen, 1982; Truscott, 1999). Such claims remain largely unsupported by empirical evidence.

In fact, researchers have amassed a wealth of empirical data demonstrating that CF facilitates L2 grammatical development. Early interaction studies found that even indications of communicative failure from an L2 learner's conversation partner leads learners to modify incorrect utterances (Lin & Hedgcock; 1996; Pica, 1988; Pica et al., 1989; Pica, Lincoln Porter, Paninos & Linnel, 1996). Other FFI studies have demonstrated pre- to post-test L2 grammatical development from conditions that included teacher-provided CF on learners' errors along with other instructional treatments (Spada & Lightbown, 1993; White, 1991; White, Spada, Lightbown, & Ranta, 1991). Additionally, studies examining CF treatments compared to non-CF treatments have reported greater progress in L2 grammatical development from pre- to post-tests for CF treatments (Doughty & Varela, 1998; Ellis et al., 2006; Mackey, 1999; Sheen, 2008; Yang & Lyster, 2010). Finally, meta-analyses of CF research provide strong support for the conclusion that CF facilitates the development of L2 grammatical knowledge (e.g., Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006).

Notwithstanding these findings, some researchers (e.g., Krashen, 1982, 1985; Schwartz, 1993; Truscott, 1999) claim that CF only allows learners to develop knowledge about the rules of a language, also known as an explicit or declarative knowledge of the language. They argue that CF cannot change learners' implicit knowledge of a language or their underlying competence (the unconscious mental representations of the language

stored in long term memory that allow learners to spontaneously produce correct utterances). No conclusive evidence has demonstrated that the role of CF is so limited. In fact, several recent studies have shown that CF treatments lead to increases in performance on tasks that may limit or even prevent learners from accessing their explicit knowledge, and instead compel them to rely upon their procedural (or near-automatic) or implicit knowledge of L2 grammar (Ellis et al., 2006; Lyster & Saito, 2010).

Research has not only demonstrated that CF facilitates L2 grammatical knowledge but also that learners' awareness of the corrective intent of CF is important in that facilitation. One problem with the provision of CF in the classroom is that it sometimes goes unnoticed by students (Chaudron, 1977b; Lyster, 1998; Nicholas et al., 2001). That is, the intent of subtly (or implicitly) provided CF can be ambiguous. For example, teachers often correct students' errors by reformulating their inaccurate utterances while maintaining a focus on meaning. This implicit form of correction, introduced in Chapter 1, is referred to as recasting, and it is thought to trigger learner-noticing, leading them to re-think the way they have just said something. However, as several researchers have pointed out, it is also possible that learners may believe that the recast is simply another way of saying the same thing, or a confirmation of the content of their utterance. This is particularly the case in communicative and content-based learning environments where the primary focus is on meaning. The question of how explicit or overt CF needs to be so that it is noticed by learners has played a large part in the CF research agenda. This line of research has primarily focused on three types of the CF found in Lyster and Ranta's (1997) often-cited CF taxonomy: recasts, prompts, and explicit correction. These three types have typically been considered to fall along a

continuum ranging from most implicit to most explicit (Ellis, 2001; Russell & Spada, 2006). The research reviewed below has revealed that while CF types that are more explicit tend to be more effective for L2 learning, more implicit CF can also be effective if that CF is given in a form or in a context in which learners are likely to be more aware of the corrective intent.

Prior to reviewing the evidence supporting this point, it is helpful to define recasts, prompts, and explicit correction. As described above, teachers recast by reformulating learners' incorrect utterances while maintaining a focus on meaning. Recasts, according to Lyster and Ranta (1997), "involve the teacher's reformulation of all or part of a student's utterance, minus the error" (p. 46), for example,

Student: The package sent to Canada.
Teacher: The package was sent to Canada.

Prompts, on the other hand, do not provide a correct model and take several different forms including the following four types:

1) Clarification requests, Lyster and Ranta explain, "indicate to students either that their utterance has been misunderstood by the teacher or that the utterance is ill-formed in some way and that a repetition or a reformulation is required" (p. 47), for example,

Student: The package sent to Canada.
Teacher: Sorry, what?

2) Metalinguistic feedback, according to Lyster and Ranta,

> contains either comments, information, or questions related to the well-formedness of the student's utterance, without explicitly providing the correct form. Metalinguistic comments generally indicate that there is an error somewhere....Metalinguistic information generally provides either some

> grammatical metalanguage that refers to the nature of the error....or a word definition in the case of lexical errors. (p. 47)

For example,

Student: The package sent to Canada.
Teacher: You need an auxiliary verb because that sentence is passive.

3) Elicitation, Lyster and Ranta explain,

> refers to at least three techniques that teachers use to directly elicit the correct form from the student. First, teachers elicit completion of their own utterance by strategically pausing to allow students to "fill in the blank" as it were ....Such "elicit completion" moves may be preceded by some metalinguistic comment such as "No, not that. It's a . . . " or by a repetition of the error....Second, teachers use questions to elicit correct forms....Third, teachers occasionally ask students to reformulate their utterance. (p. 48)

For example,

Student: The package sent to Canada.
Teacher: Try that sentence once more.

4) Repetition, Lyster and Ranta note, "refers to the teacher's repetition, in isolation, of the student's erroneous utterance" (p. 48), for example,

Student: The package sent to Canada
Teacher: The package *sent*?

Finally, explicit correction is defined by Lyster and Ranta (1997) as "the explicit provision of the correct form. As the teacher provides the correct form, he or she clearly indicates that what the student had said was incorrect" (p. 46), for example,

Student: The package sent to Canada
Teacher: That's not right. You have to say, the package was sent to Canada

The following review of research conducted on these CF types reveals that the effectiveness of CF is impacted by how aware learners are of the corrective intent of the CF.

2.2.1. Recasts and prompts.

Amongst CF types, recasts have attracted the greatest amount of research interest. They are repeatedly found to be the most commonly employed type of CF by teachers (Ellis et al., 2001; Loewen & Philp, 2004; Lyster & Mori, 2006; Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2004; Yoshida, 2008). Nonetheless, as noted above, the implicitness of recasts is often blamed for learners failing to appreciate their corrective intent (e.g., Chaudron, 1977b; Fanselow, 1977; Lyster, 1998; Nicholas et al. 2001). Evidence that learners often miss the corrective intent of recasts is most apparent in studies that compare recasts and prompts. To compare the effectiveness of these CF types, researchers have 1) compared measurements of learner uptake immediately after CF has been provided and 2) compared scores on tests that have followed instructional treatments which have featured CF.

Lyster and Ranta (1997) introduced the concept of uptake (Austin, 1962) as "a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance" (p. 49). However, the illocutionary force of recasting does not compel uptake like the force of prompting does. Furthermore, given that recasts are often interpreted as confirmations of preceding utterances, learners and teachers often appropriately continue the conversation, thus eliminating the chance for learners to indicate uptake. By excising such occasions in her analysis of the effect of recasts, Oliver

(1995) observed a noticeable increase in the number of instances of uptake after recasts. Additionally, Loewen and Philp (2006) found that although uptake from a prompt-treatment group was higher than that of a recast-treatment group, there were no statistically significant differences in the scores of the two groups on a post-test of grammatical accuracy. Moreover, Mackey and Philp (1998) found that even though a group of low proficiency learners displayed an equally low rate of uptake on recasts compared with a group of higher proficiency learners, post test results indicated statistically significant advantages over the longer term for the higher proficiency group. The findings from these three studies demonstrate that any simple comparison of the amount of uptake from prompts to the amount of uptake from recasts is not a reliable way of determining which type of CF is better at facilitating SLA.

Uptake has also been examined to determine how aware learners are of the corrective intent of recasts. For example, Mackey, Gass, and McDonough (2000) reported learners' failure to produce uptake after recasts that were provided in communicative activities. Afterward, the researchers asked learners to comment on the recordings of their own production. They found that learners who failed to display uptake after recasts also failed to indicate awareness that in those instances of recasts, their grammar was being corrected.

Analyzing uptake has also revealed different uptake patterns in relation to CF. For example, recasts have been found to lead to greater uptake when they are short, focus on one error, feature repetition, or appear in the form of statements rather than questions (Chaudron, 1977b; Lyster, 1998; Sheen, 2006). However, no statistically significant difference has been found in the amount of uptake resulting from prompts and recasts in

studies that feature recasts enhanced in some manner to make them more noticeable (Nassaji, 2007; O'Relly, Flaitz & Kromrey, 2001; Sheen, 2004). Furthermore, in studies that report no attention-drawing modifications to recasts, they regularly result in significantly lower percentages of uptake than prompts (Ellis et al., 2001; Lyster & Ranta, 1997; Morris, 2002; Panova & Lyster 2002; Yang & Lyster, 2010). Thus, when the form of the recast makes the corrective intent more obvious, learner uptake increases, but when the form of the recast is not manipulated, learners often miss the corrective intent of the recasts and produce a lower amount of uptake.

It has also been revealed that context plays a role in learners' awareness of the corrective intent of recasts and correspondingly in the amount of uptake that results from recasts. Learners in certain contexts appear to be more aware of the corrective intent of recasts than in others. That is, recasts tend to lead to low amounts of uptake in instructional contexts that are strongly oriented to communication (e.g., French immersion and other content-based instruction classes) rather than language (Lyster & Ranta, 1997; Morris, 2002; Panova & Lyster, 2002). However, in contexts where the instruction is more form-oriented, the uptake from recasts greatly increases (Ellis et al., 2001; Sheen, 2004). Moreover, in studies that have compared uptake rates in relation to recasts and prompts, recasts usually result in a lower amount of uptake than prompts (Lyster & Ranta, 1997; Morris, 2002; Panova & Lyster 2002). The exceptions to this pattern occur in contexts where instruction is form-oriented (Lyster & Mori, 2006; O'Relly et al., 2001; Sheen, 2004)

In summary, when the corrective intent of recasts is present, and/or they are provided in contexts where the instruction is more oriented to form and accuracy, there is

a greater chance for the corrective intent of recasts to be noticed and more uptake may be displayed. Conversely, when recasts do not overtly signal corrective intent or they are provided in contexts where learners are focused on the communication of messages and meanings, learners are more likely to miss the corrective intent and less uptake is likely to be displayed.

Such inconsistency in learner awareness of the corrective intent of recasts has been suggested as one explanation for the mixed results of studies attempting to demonstrate the effectiveness of recasts in facilitating SLA (Doughty, 2001; Nicholas et al., 2001). This has led to the argument that recasts need to be enhanced in some way, or be provided in instructional contexts where attention to form is emphasized in order to make learners aware of their corrective intent. Empirical evidence supports this argument. For example, Doughty and Varela (1998), one of the most cited classroom studies of the benefits of recasts for L2 development employed what they referred to as 'corrective recasts'. These recasts were explicit because they consisted of two moves: repetition of learners' errors with emphasis followed by a reformulation (i.e., correct model). When comparing the effects of 'corrective recasts' with no corrective feedback, the researchers found clear advantages for 'corrective recasts'. In subsequent research, however, when the effects of prompts and recasts (i.e., not 'corrective recasts') on L2 learning have been compared, prompts regularly lead to significantly better performances on post tests of grammatical accuracy than recasts (Ammar & Spada, 2006; Ellis, 2007a; Ellis et al., 2006; Lyster, 2004; Yang & Lyster, 2010). Furthermore, Lyster and Saito's (2010) meta-analysis of classroom CF studies found a larger effect size for prompts than recasts.

However, there are exceptions to this pattern of prompt superiority.

Unsurprisingly these exceptions come from studies in which learners are in more form-oriented contexts (such as laboratories), and therefore are more likely to be sensitized to the corrective intent of recasts. For example, Lyster and Izquierdo (2009) found no difference between the effectiveness of prompts and recasts in a laboratory study. Also, Mackey and Goo's (2007) CF meta-analysis confirmed that recasts are more facilitative of learning and learner noticing in laboratory settings than in classroom settings. Thus, there is evidence that without some emphasis in the presentation of recasts or when recasts are provided in meaning- rather than form-focused contexts, their corrective intent can be missed by learners, and therefore that awareness of corrective intent plays an important role in the effectiveness of CF in facilitating the development of L2 grammar knowledge.

2.2.2. Explicit correction.

Results from explicit correction studies also indicate that the learners' awareness of the corrective intent of CF plays a beneficial role in facilitating the development of L2 grammar knowledge. In a series of explicit correction studies known as 'the garden path studies,' Tomasello and Herron (1988, 1989) found that inducing learners into making errors, and then providing the correct model after those errors demonstrated that explicit correction facilitates improvement. Furthermore, in comparative studies, explicit correction has proven more effective than recasts (Carroll & Swain, 1993; Ellis et al., 2006; Sheen 2007). These findings add more support to the position that CF serves learners better when they are made more aware of its corrective intent.

2.2.3. Summary.

Research exploring the benefits of different types of CF has demonstrated that it is

most effective in the development of L2 grammatical knowledge when learners are aware of the corrective intent of the CF. These findings come from research in which CF has been provided immediately. As argued above, delayed CF is likely to be equally or more effective than immediate CF because the corrective intent of delayed CF is necessarily more explicit leaving little room for learners to misinterpret or fail to notice it.

2.3. Theoretical Frameworks for Immediate and Delayed CF

The evidence that CF is most effective when learners are aware of its corrective intent is in line with Schmidt's (1990, 1995) noticing hypothesis which states that learners need to notice features of an L2 as a first step towards acquiring them. As just explained, unlike some types of CF that are provided within communicative practice, such as recasts, there is almost no risk that learners will fail to notice the corrective intent of delayed CF. Thus, it is reasonable to hypothesize that delayed CF would facilitate the development of L2 grammatical knowledge. However, as is discussed below, the frameworks that traditionally explain why immediate CF facilitates the development of L2 grammar knowledge do not provide an explanation for why delaying CF would also do so. Thus, it is important to explore what theoretical justification there is for claiming that delayed CF facilitates the development of L2 grammatical knowledge.

2.3.1. Traditional theoretical frameworks for CF.

Skill acquisition theory (SAT) and immediate cognitive comparison (ICC) are two theoretical frameworks that are frequently referred to in attempts to explain why CF provided during communicative practice facilitates the development of L2 grammatical knowledge. As will become clear, neither framework, *as it has thus far been interpreted by researchers*, explains why delayed CF might be useful in facilitating the development

of L2 grammatical knowledge.

2.3.1.1. Skill acquisition theory.

SAT is a cognitive psychology theory that posits that humans learn to do things, such as drive, by first learning factual information about a skill and then practicing until the skill becomes automatic. DeKeyser (1998, 2001, 2007) argues that a version of SAT called Adaptive Control of Thought (Anderson, 1982, 1983, 1993) can be applied to SLA. Doing so requires an initial explanation about a language form, such as providing rules about a grammar point. Then, exercises ensure that learners gain factual (declarative) knowledge of those rules. The form must then be proceduralized, or learned as a behaviour. Learners must practice the linguistic behaviour in a meaningful context while keeping the declarative knowledge in mind. When errors are made, correction is required to help learners realize what they have done wrong. After the declarative knowledge is converted into a proceduralized behaviour, it must be practiced many times in meaningful situations until it becomes automatic.

In an experiment demonstrating that SAT could be applied to SLA, DeKeyser (1997) included a proceduralization stage in which learners' attention was drawn to form as they did communicative drills which forced them to try to convey meaning while keeping the declarative rules of what they were saying in mind. Computerized corrective feedback drew their attention to failure in producing the correct behaviour. Ranta and Lyster (2007) argue that CF in the form of prompts may facilitate proceduralization in communicative practice because prompts encourage learners to retrieve the information that they have about a grammar rule and try to produce the language again more accurately with that declarative knowledge of the rule in mind. In other words, they argue

that prompts instantiate what DeKeyser (1997, 2007) describes as proceduralization. DeKeyser (2007) agrees that CF in the form of prompts can facilitate the proceduralization of knowledge because prompts compel learners to practice producing L2 forms which they have not yet mastered. However, because prompts in the instructed SLA research literature are always given immediately after errors during, and not after, communicative activity, it is not obvious how the current interpretation of SAT as applied to SLA could be used as a theoretical framework to explain any effectiveness delayed CF might have in facilitating the development of L2 knowledge.

2.3.1.2. Immediate cognitive comparison.

The second traditional explanation for the effectiveness of CF in facilitating the development of L2 knowledge is ICC. In ICC, CF comes in the form of a model of a correct version of an utterance which is immediately compared with a learner's error and is intended to make the learner notice the difference between what they have produced and the correct way to produce it. Ranta and Lyster (2007) labeled this kind of CF reformulation. Recasts and explicit correction are reformulations because they always re-work learners' errors into corrected models unlike prompts that always withhold correct responses. Ellis (2006) characterized the former as CF that pushes output and the latter as CF that provides input.

Reformulations are hypothesized to facilitate SLA by providing accurate models as input that learners compare to the errors that they have just committed while both are still fresh in the memory of the learner (Ellis, 2006; Ranta & Lyster, 2007). Importantly, ICC offers no obvious theoretical support for delayed corrective feedback due to what Doughty (2001) calls "the cognitive window". This window refers to the approximately

40 seconds that humans can maintain active mental representations in working memory. According to Cowan (1995), working memory is the activated component of long-term memory that facilitates the processing and storage of new sensory input. Since the comparison in delayed CF happens after a task has been completed, the delay between the error commission and the corrected model to be compared to it would likely exceed one minute. This amount of time would be too long to retain the error commission in working memory, leaving nothing for the learner to compare to the teacher-supplied corrected model of CF.

Clearly there are differences in the theoretical positions that support immediate corrective feedback. Nonetheless, these diverse theoretical perspectives share the assumption that CF is best provided within the context of communicative interaction. In other words, the current theoretical frameworks discussed in CF research only support immediately provided CF.

2.3.2. Alternative theoretical frameworks for immediate and delayed CF.

Any grammatical development found to result from delayed CF requires an explanation that goes beyond what has been offered until now in the theoretical literature about immediately provided CF. In what follows, I suggest that need may be potentially addressed by one or more of three theoretical frameworks: transfer appropriate processing (TAP), preparatory attention, and reactivation and reconsolidation (RAR). Each of these frameworks provide explanations for both immediate and delayed CF. As such, they increase the breadth of the theoretical foundations upon which research into the timing of CF can proceed.

2.3.2.1. Transfer appropriate processing.

The TAP framework posits that memories are best recalled in conditions similar to those in which they were encoded (Morris, Bransford, & Franks, 1977). A modified version of an analogy used in Morris et al. (1977) is helpful in explaining this concept. Imagine two groups of participants were exposed to the same list of words. Group 1 was instructed to check for vowels at the end of each word, and group 2 was told to consider whether each word was a plant, animal or inanimate object. Imagine that a post-test consisted of all students estimating the number of words in the list that ended in "e." In such a case, TAP predicts that group 1 would perform faster and more accurately than group 2 because the test focuses on the structure of the word, and group 1 was instructed to process the words from the original list for a structural property. Similarly, imagine there was a different post-test that asked all students to tell if a subsequent list of words were synonyms to words on the list. In that case, TAP predicts that group 2 would out-perform group 1 because the second test is a test of meaning, and group 2 was instructed to process the original list for meaning. Thus, TAP predicts that humans retrieve memories best when we retrieve them in the way we originally processed them.

Some SLA researchers have suggested that TAP might have an impact on instructed SLA (e.g., Lightbown, 2008; Segalowitz & Lightbown, 1999; Spada & Lightbown, 2008). Spada and Lightbown (2008) suggest that learners who are instructed on grammar structures during communicative practice (i.e., integrated form-focused instruction) might score significantly better on a test of those structures that takes the form of a communicative task then on a discrete point grammar test. Concomitantly, different learners who learned the same grammar feature by learning the rules of the

feature outside of a communicative task (i.e., isolated form-focused instruction) might score higher on a discrete grammar point test than on a test in the form of a communicative task. Another way of looking at this interpretation of TAP is that providing grammar instruction that is integrated into communicative activity may lead to an increase in procedural (potentially even implicit) grammar knowledge while providing isolated grammar instruction may lead to an increase in explicit (i.e., declarative) knowledge.

What would be the result of extending this interpretation of TAP to the potential effects of immediate and delayed CF? Because immediate CF is provided during a communicative activity, it is arguably integrated grammar instruction. Conversely, delayed CF is provided after communication and as such should arguably be considered isolated grammar instruction. Presumably, if Spada and Lightbown's interpretation of TAP were extended to immediate and delayed CF, delayed CF should lead to better results on tests of explicit (i.e., declarative) knowledge than immediate CF does, but immediate CF should result in better results in tests of procedural (i.e., implicit) knowledge. As such, delayed CF would be predicted to be constrained to facilitating an increase in development in only explicit grammatical knowledge. Such a prediction is in line with Schwartz (1993) and Krashen's (1982, 1985) position on CF discussed above. In any case, TAP predicts that L2 development (albeit differential L2 development) could result from both immediate CF and delayed CF.

2.3.2.2. Preparatory attention.

Another potential explanatory theoretical framework is called preparatory attention. This theory claims that when difficult tasks force learners to realize that they

need more information about how an L2 works, the learners may pay more attention to any new information that subsequently is made available to them. That is, the trouble learners have with the language sensitizes them to be more attentive to further instruction. This sensitization is called preparatory attention (LaBerge, Auclair, & Seiroff , 2000; N. Ellis, 2005).

Hondo (2009) suggested that preparatory attention may have facilitated the positive results of delayed instruction in her study. She found that on comprehension tests for an English modal verb, learners given delayed attention to form significantly outperformed learners who had received instruction before communicative practice. She argued that delaying instruction until 10 minutes into communicative practice might have fostered preparatory attention in learners. Hondo operationalized the delay through instruction, rather than CF, and the delay never extended outside of communicative practice. Nonetheless, Hondo's results demonstrated that attention to form need not necessarily be provided immediately at the moment that learners notice a gap between what they know and what they do not know. Hondo's finding suggests that attention to form can be delayed and remain effective at least when that attention to form is preceded by challenging material that may induce preparatory attention. Arguably, delayed CF which followed material that induced preparatory attention would also be effective. Thus, preparatory attention may provide theoretical support for why delayed CF might facilitate the development of L2 grammar knowledge. Preparatory attention could also explain development that resulted from immediate CF. Learners engaged in a communicative task who encounter trouble and are given immediate CF presumably become more sensitized to subsequent provisions of immediate CF.

2.3.2.3. Reactivation and reconsolidation.

RAR is yet another potential theoretical explanation for why not only immediate but also delayed CF may facilitate the development of L2 grammar knowledge. RAR is an area of recent research on memory in cognitive psychology. Essentially, this research demonstrates that simply retrieving a memory temporarily returns that memory to a similar state of plasticity (i.e., makes that memory labile or susceptible to change) as to when it was originally formed. When humans recall something, the memory is *reactivated* into a state that allows for new information to alter or update the memory before the brain *reconsolidates* and stores the memory again. RAR has been demonstrated to occur in both procedural and declarative memory, and it has recently been used in a bilingualism study as an explanation for an L2-L1 transfer phenomenon.

It is important to differentiate the memory *reconsolidation* in RAR from memory *consolidation*. Memory consolidation is the initial process of memory storage. When the input that will eventually become a memory first enters the mind, it is in a fragile state. That is, it does not represent a fully-formed memory. After a day of consolidation, this knowledge becomes much less fragile. Until the inception of RAR research, it was assumed that as time progressed, newly-consolidated memories continued to consolidate, becoming less and less susceptible to change (Hupbach, Gomez & Nadel, 2009; Nader and Einarsson, 2010). RAR research has demonstrated that long term memories are, in fact, susceptible to change under the right circumstances and that when they are reactivated and made susceptible to change, the mind attempts to reconsolidate them. Often, however, old ideas which have temporarily been made labile again through the

process of retrieval are altered when they are exposed to new information, and that information is incorporated with the old idea during the reconsolidation process.

Walker, Brakefield, Hobson, and Stickgold (2003) found that RAR affects procedural memories. These researchers trained participants to finger tap two different sequences using keyboard keys labeled 1 to 4. Sequence A was 4-1-3-2-4 and sequence B was 2-3-1-4-2. Although 8 groups were included in the study, the relevant findings for this literature review involve only two groups, henceforth referred to as group 1 and group 2. For both groups, on day 1, sequence A was learned; on day 2, sequence B was learned; and on day 3, sequence A was tested. There was one essential difference in the process for group 1 and group 2. On day 2, immediately before group 2 learned sequence B, they rehearsed sequence A. Group 1 did not do so. On day 3, group 1 suffered no interference effects in their performance of sequence A. Group 2, however, suffered statistically significant interference from sequence B that resulted in decreases in speed and accuracy due to the intrusions from sequence B. Walker et al. (2003) concluded that when group 2 rehearsed sequence A, they reactivated the procedural memory for sequence A, thus returning it to a labile state which needed to be reconsolidated. When sequence B was introduced shortly afterward, the similarity of sequence B to sequence A caused sequence B to intrude and interfere with the reconsolidation process for sequence A. The results of the interference in reconsolidation were present in the test of sequence A on day 3.

Forcato, Burgos, Argibay, Molina, Pedreira, and Maldonado (2007) and Hupbach, Gomez, Nadel, and Hardt (2007) demonstrated that declarative memory is also subject to RAR. These researchers used similar 3-day experimental designs as Walker et al. (2003)

Instead of proceduralizing a motor skill, the learners in this research were all exposed to items to be stored in their declarative memories. For Forcato et al. (2007), the participants were taught lists of associated words, and Hupbach et al.'s (2007) participants remembered physical objects. In these studies, on days 1 and 3 both groups of participants (i.e., reminded and non- reminded conditions) were exposed to the items, and on day 2 participants were introduced to a separate group of similar items. Again, day 2 featured a difference in treatment for the reminded and non-reminded conditions. On day 2, the researchers elicited memories of the items from day 1 from the participants in the reminded condition; participants in the non-reminded condition were given no reminders. The findings in these studies were very similar to those from Walker et al. (2003). Only the reminded condition showed a statistically significant number of intrusions from the items presented in day 2. Because of the asymmetry in the direction of the intrusions (i.e., day 2 items led to statistically significant confusion with day 1 items and not the other way around), the researchers concluded that the reminders led to reactivation of the declarative memories that were returned to a state of lability, which allowed for the memories to be updated by the additional information provided on day 2.

To my knowledge, Wolff and Ventura (2009), in their study of L2-L1 transfer, were the first (and thus far only) researchers to apply the RAR framework (or as they label it, "retrieval-induced reconsolidation," p. 153) to applied linguistics. They compared monolinguals and bilinguals' use of verbs used for describing "causing and enabling" in descriptions of some animations. They found that descriptions significantly differed between monolinguals and bilinguals. Even though bilinguals used their L1s to describe the animations, their descriptions followed the conceptual pattern more typical

of their L2. Thus, bilinguals of Russian and English who spoke Russian as an L1 used Russian words but an English conceptual structure to describe the animations. Bilinguals of English and Russian who spoke English as an L1 similarly used English words but a Russian conceptual structure to describe the animations. Monolinguals of Russian and English used the words and conceptual styles of their own languages. Wolf and Ventura interpreted these results to indicate that while learning their L2, bilinguals reactivate L1 analogues causing these analogues to be labile and susceptible to intrusion from the L2. The resulting changes are evidenced in the L2-L1 transfer.

The RAR process has been demonstrated to affect both procedural and declarative memory and has even been used to explain the influence from, and alteration of, linguistic long term memory representations. Though Wolff and Ventura (2009) only apply the RAR framework to L2-L1 transfer, RAR seems well suited as a theoretical explanation for why CF leads to grammatical development. The experimental operationalization of RAR in both the procedural and declarative memory studies above shares qualities with prompts and reformulations. For example, when a teacher prompts a learner to correct a malformed utterance, the learner must attempt to produce the correction, and doing so entails retrieval of their mental representation of how to form that utterance. In RAR, it is the retrieval of a mental representation that allows for a subsequent stimulus to alter the original representation. Reformulations also share similarities with the RAR process. For example, in the RAR studies, one mental representation is hypothesized to lead to the alteration of another, just as teacher-reformulated models lead to alterations of incorrectly formed utterances.

How could teachers operationalize CF that would facilitate restructuring of incorrect linguistic long term memory representations through RAR? To explore this question, I will briefly summarize how RAR was operationalized in the procedural and declarative memory studies above. First, participants were given a procedural or declarative stimulus which began to consolidate into a long term memory representation in their minds. Second, after that stimulus had been consolidated further, the participants were reminded of it, thus returning the long term memory representation to a labile state. Third, the participants were exposed to a stimulus which was slightly different from the now-labile long term memory representation. When reconsolidation of the original long term memory representation occurred, features of the interference stimulus were made part of the original long term memory.

How might one extrapolate RAR theory to questions about the provision of delayed and immediate CF in response to learner errors? Consider the following scenario: the teacher asks learners to complete a communicative task that compels them to use a particular language form. When learners produce the feature incorrectly, the teacher concludes that their knowledge of that feature is not fully developed. Put differently, teachers would be able to see that the learners have a long term memory representation of how to form the feature which does not yet match the long term memory representation of the native speakers of that language. In the case of immediate CF, the teacher could prompt the learner to try again and then provide a model and ask the learner to produce that model. In this way, the original incorrect mental representation will have been retrieved and made labile in close proximity to a second representation which is similar to the original but accurate. The latter representation should intrude into the original

inaccurate representation, causing parts (if not all) of the correction to become part of the original mental representation when the original representation is reconsolidated. As this process is likely to be fairly obtrusive, it may facilitate L2 development through immediately provided CF, but it is probably better suited for CF which is delayed. In that case, after a communicative activity has revealed an emergent but inaccurate mental representation of a feature, teachers could wait until the activity is complete before prompting the learner to attempt the feature once again. Once the mental representation has been retrieved, the same process as has just been described, could be followed.

2.3.3. Summary.

CF research to date has almost invariably focused on the effects of CF that are provided immediately after learner error, and there is theoretical support for such practice as articulated in theories of skill acquisition and immediate cognitive comparison. There is also theoretical support for the provision of immediate and delayed corrective feedback informed by theories of transfer appropriate processing, preparatory attention, and reactivation and reconsolidation. Nonetheless, as is evident from the review that follows, only a limited amount of instructed SLA research has included the investigation of delayed CF despite the fact that L2 teachers are often encouraged to provide delayed CF in their instructional practice.

2.4. Research on Delayed CF

Research that has focused primarily on the timing of CF can be divided into four types. First, there are descriptive studies that focus on the methods used to provide delayed CF in L2 classes. Second, there is research that has focused on the L2 development that has resulted from delayed CF on written work, which can be considered

analogous to delayed CF on spoken errors. Third, there is research that focuses on the developmental effectiveness of delayed CF. Finally, there are a few studies that compare the developmental effectiveness of delayed and immediate CF. The limitations of this research are highlighted to demonstrate that further research is required into the developmental effects of CF that is provided at different junctures in time.

2.4.1. Descriptive research on delayed CF.

To my knowledge, only Rolan-Ianziti (2006, 2010) and Hunter (2012) have conducted observation studies with the central aim of describing teachers' use of delayed CF. Rolin-Ianziti (2006) focused on how delayed CF was being used in a French as a foreign language class in Australia. Teachers in this course felt that the value of role plays as opportunities to practice communicative use of French would have been stifled by interruptive immediate CF, so they wrote down learners' errors and delayed CF until the end of role plays. The 2.5 hours of observation in the study consisted of three teachers giving delayed CF to 15 groups of two to three students each. Rolin-Ianziti found that teachers addressed linguistic errors after the role play ended in two ways: teacher reviewing and teacher initiating. These two processes were essentially delayed versions of explicit CF and prompting respectively. In the qualitative examples that Rolin-Ianziti provides, teachers give CF to individual students in front of the class, and the CF focuses on both pronunciation and grammar errors. In teacher reviewing, teachers explained what the error was, and provided correct models. In teacher initiations, teachers asked students how to perform linguistic functions that had been part of the role play such as asking students how to ask someone's name or profession. Rolin-Ianziti did not provide a statistical analysis, but she did report that teachers used similar amounts of reviewing and

initiation. Furthermore, she indicated that initiation was superior to reviewing in terms of both uptake and repair.[5]

Rolin-Ianziti (2010) used conversation analysis (Schegloff, Jefferson & Sacks, 1977) to perform a microanalysis of a sub-set of the data from her 2006 study described above. She re-labels the teacher-reviewed and teacher-initiated correction types teacher-initiated correction and teacher-initiated student-correction respectively. Rolin-Ianziti's (2010) in-depth analysis of the latter type of CF reveals that this form of delayed prompting is executed through the following "initiators": category questions, designedly incorrect utterances, and requests to quote from the role play. For example, if a learner made an error such as, "I are a bus driver", then, after the communicative activity was complete, the teacher had three choices of initiators with which to prompt. The teacher could use a category question such as, "Can you tell me again what your job in the role play was?" Alternatively, the teacher could prompt with a designedly incomplete utterance, saying first, "I are a bus driver?" "You should say, I..." and wait for the student to attempt the correct construction. The third option was to request a quote saying, for example, "Alright, so did you say your job? You did? What did you say again?" Regardless of whether or not the student is able to self-correct accurately, the teacher eventually provides the accurate model at the end of the teacher-initiated student-correction and encourages the student to repeat the model.

This research was not motivated by any particular theoretical framework. However, the type of delayed CF used by the teachers in Rolin-Ianziti's studies provide potential models for how one might operationalize delayed CF based on RAR theory. For example, the teacher elicits retrieval of the mental representation of how the learner

[5] Repair is uptake in which learners accurately self-correct (Ellis et al., 2001; Lyster & Ranta, 1997)

thinks a feature is supposed to be formed and then quickly provides a correct model which the learner is urged to rehearse. As indicated above, that model should become a mental representation which will intrude into the reconsolidation of the originally malformed mental representation. Rolin-Ianziti (2006, 2010) makes no claims about whether or not delaying CF facilitates the development of L2 grammar knowledge in either of her studies, but Rolin-Ianziti (2006) calls for investigation into the potential developmental effects of delayed CF.

Hunter (2012) provides delayed CF through an innovative teaching method called "Small Talk" (Harris, 1998). "Small Talk" involves small groups of students completing short discussion tasks while their teacher monitors and keeps a record of errors from individual students. This information is used to compile computerized student worksheets that list students' personal errors as well as some errors from classmates. Within 24 hours of the end of a class, students can download the worksheets and accompanying audio files that feature accurate models of the errors. First, students attempt to correct the errors on their own with the helpful metalinguistic hints that are provided on the worksheets, and/or second, they can use the recorded models to assist them in making the correction. Running lists of errors continue to develop as the semester proceeds, and students are given regular running-list tests in which they must correct the errors from their running list under time pressure.

Hunter (2012) reported on the results of a small-scale study of "Small-Talk" consisting of 12 intermediate-level adult English for Academic Purposes students in four "Small-Talk" sessions. He found that teachers provided CF for an average of 40% of the participants' errors. Hunter concluded that this finding meant that 40% of students' errors

would eventually result in repair because students would subsequently be obligated to correct their errors on their worksheets. He noted that this compared favourably with Lyster and Ranta's (1997) finding that only 17% of errors would result in repair from the immediately provided CF observed in their study.

As this was a descriptive study, Hunter (2012), like Rolin-Ianziti (2006, 2010) made no developmental claims about the results of the delayed CF, but rather indicated that "Small Talk" provides a method that allows teachers to provide delayed CF on oral production without interrupting communicative activity. Despite the fact that spoken models are provided in audio files, some might suggest that the inclusion of written versions of the errors and written metalinguistic hints on the worksheets make Hunter's delayed CF seem more like written delayed CF.

2.4.2. Delayed CF on written production.

Researchers' calls for investigations into whether delayed CF facilitates SLA as effectively as immediately CF are understandable not only because delayed CF takes place in classrooms, but also because an analogous form of delayed CF on oral production – delayed CF on written production – has been found to facilitate L2 development. Unlike CF on oral production, CF on written work is usually delayed because typically teachers need time to mark written texts before returning them to learners.

In his critical reviews of written CF Truscott (1996, 1999) indicated, methodological limitations, such as the lack of control groups, weakened the case that written CF could facilitate the development of L2 grammar knowledge. However, more recent studies designed to address those criticisms have shown that written CF can

facilitate the development of L2 grammatical knowledge (e.g., Sheen, 2007; Sheen, Wright & Moldawa, 2009; Sheen, 2010a, Van Beuningen, De Jong & Kuiken, 2012).

The growing evidence that written delayed CF is effective in facilitating L2 development does not mean that delayed oral CF also facilitates L2 development; writing and speaking are different modalities. Written CF provides a permanent record that learners can reflect upon whenever and as often as they wish to do so, whereas oral CF is much more ephemeral (Williams, 2012). Even though Sheen (2010a) found no modality-based differences in her comparison of written and oral metalinguistic CF, she did find that written reformulation treatments significantly outperformed orally-provided recasts. Furthermore, as Bitchener (2012) suggests, the amount of time that written CF allows learners to reflect on and respond to CF may allow for deeper processing, and may even make written CF more effective than oral CF. At the very least, it would be unwise to assume that the same developmental outcomes that result from delayed written CF are guaranteed to result from delayed oral CF.

2.4.3. Developmental research on delayed CF.

The preceding review of descriptive research on oral CF and experimental written delayed CF research points to the need for experimental investigation into the developmental effectiveness of providing delayed CF on spoken errors. Hunter's (2011) unpublished doctoral dissertation represents one attempt at addressing this need. Hunter (2011) is a study on the developmental results of the "Small Talk" method that was described above. Twenty-six students from two intact classes of L2 learners participated in 14-week English programs that included "Small Talk." Results from the running list tests, described above, indicated that the participants were performing at an average of

above 70% accuracy. Hunter concedes that critics might argue that successful repair on such tests does not entail full acquisition. Furthermore, he also concedes that such successful results may be indicative of the learning of items rather than systematic acquisition of grammar structures.

However, beyond these conceded limitations, some might also question whether the type of delayed CF used in the "Small Talk" method should actually be considered delayed CF because the delayed CF that is part of "Small Talk" is provided on a different day than the lesson in which the errors were made. Such CF might at the very least more appropriately be labeled "postponed CF" in keeping with Chaudron's (1977b) original categorization. Arguably, a "delay" in common parlance refers to a small deferral in time, while a "postponement" suggests that some event is being re-scheduled on a subsequent day. Finally, some might argue that treating errors on a subsequent day should not even be categorized as CF, but instead should be considered as remedial instruction based on the diagnosis of previously observed weaknesses. Such a perspective is persuasive in the case of Hunter (2011, 2012) for two reasons. First, not all of the errors on students' worksheets are their own. In other words, the worksheets force students to address errors from other learners that may or may not be shared by the individual student who is working on them (although, see Ohta, 2000, for a CF study in which it is argued that learners benefit from CF provided on classmates' errors.) Second, the more detailed description of how errors are treated in the "Small Talk" method (Hunter, 2011) suggests that the error treatment is more reminiscent of remedial instruction than CF:

> A variety of activities are used to focus attention on the CF provided by worksheets, from quick warm-up activities to focused grammar instruction....The

> students periodically practise and are tested on their own worksheet items in class, to promote automatic use of the targeted forms. Each student thus keeps a 'Running List' of the errors (with no other markings), and practises these in class with other students, for example by giving a copy to a partner and saying the corrected versions to see if the partner can hear the differences. (Hunter, 2011, p. 70-71).

These details present a conundrum. Either it becomes impossible to claim what development results from the CF as distinct from the other instructional activities, or these activities are meant to be viewed as components of the delayed CF. In any case, there is a need for more discussion about whether postponed CF should be considered to be CF or remedial instruction. The issue of how to categorize CF that is provided at different junctures in time is addressed in more detail in Chapter 5.

Hunter's (2011) study does suggest that CF, or some form of error treatment, which is not immediate can be helpful, but the limitations of this study indicate the need for more research to be conducted before a definitive determination can be made about the effectiveness of delayed CF in facilitating L2 development.

2.4.4. Developmental research comparing delayed CF to immediate CF.

To my knowledge, there are no published studies that compare the developmental effects of delayed CF to those of immediate CF. In fact, even within unpublished research, there appear to be only three studies that address this issue: Sheen's (2012) Second Language Research Forum Conference presentation, Siyyari's (2005) MA thesis, and Varnosfadrani's (2006) PhD dissertation.

Sheen (2012) conducted a comparative study of the L2 development that resulted

from immediate and delayed metalinguistic CF on adult ESL learners' acquisition of the past tense. Learners benefited from both immediate and delayed CF. Unfortunately, the sole source of information about this study and its results is a conference abstract.[6] As such, little is known about the type of treatments and language measures that were used. Moreover, there is no information provided as to whether a control group was included in the study.

Siyyari (2005) also conducted a comparative study of the developmental effects of immediate and delayed CF. Over a period of 12 lessons, Siyyari had 4 teachers provide immediate CF to 30 adult Iranian EFL students and delayed CF to 30 others as they completed dictogloss text-reconstruction tasks. These tasks were designed to elicit the production of four grammar forms: "I wish", three forms of causative clause, second-conditional sentences, and "should have plus past participle". Three dictogloss tasks were created for each structure. The study used a pre- to post-test design. The post-test was administered 20 days after the treatment, and it consisted of explicit questions that participants were asked to respond to by completing sentences that forced the usage of the structures under investigation. For example, in testing for participants' ability to use the form "should have plus past participle", the participant would read, "I did not listen to my parents when I was younger, and for many times I had many problems because of that". Then, the participant was instructed to complete the following sentence in writing: "I should/shouldn't..." Participants improved significantly on all structures in both immediate and delayed CF conditions, and Siyyari interpreted this to mean that both immediate and delayed CF resulted in L2 grammatical development. No statistically significant differences were found to result from the immediate and delayed CF

[6] No report of the study has yet been published.

treatments for each structure. However, when an aggregate of all the scores for all structures was calculated, the Immediate CF treatment was found to have led to a significantly greater improvement than the delayed CF treatment.

One limitation of this study was that in the dictogloss tasks, the participants reconstructed the texts by using written notes they had taken when they initially read the text. Thus, it is debatable whether the CF in this study should be considered to have been provided on written or oral production. A more serious methodological problem was that recasts were provided as the immediate CF treatment while explicit CF was provided as the delayed CF treatment. Therefore, the results cannot confidently be attributed to a manipulation in the timing of CF because it is equally likely that the difference in explicitness in the two conditions could have led to the results. Furthermore, the study lacked a non-CF control group, so it is not possible to determine whether improvement came from the CF treatment or from something else in the instruction. Finally, Siyyari's method of assessing development limits the confidence one can have about the results because, as illustrated above, the language measures amounted to little more than written versions of mechanical drills.

Varnosfadrani (2006) ambitiously compared not only the development effects of delayed versus immediate CF, but also explicit versus implicit CF, CF on earlier versus CF on later acquired features, and CF on morphology versus CF on syntax aspects. Only the comparison between delayed versus immediate CF is relevant for this review. Varnosfadrani's 28 adult Iranian EFL participants completed two text reconstruction tasks. During the first task, learners received explicit immediate CF. During the second task, the same participants received explicit delayed CF. Tailor-made tests (LaPierre,

1994) were created for each error that was corrected.[7] Statistically significant improvement was found on the errors that had been made, but there were no statistically significant differences between the improvements that resulted from immediate or delayed CF.

One important limitation to Varnosfadrani's (2006) study is that, as was the case for Siyyari (2005), CF type appears to have been confounded with CF timing which prevents a true comparison of the effects of timing on CF. Varnosfadrani reports that he sometimes repeated the students' error when providing delayed CF but that immediate CF simply followed the learners' error. Thus, there may have been a comparison at least some of the time between two different types of CF rather than just a comparison of immediate and delayed CF.

Limitations to the study also resulted from the use of tailor-made tests. Researchers cannot provide pre-tests for tailor-made tests because such tests are created after activities are finished, and they are populated with items that are based on errors that learners made during the activities. The lack of a pre-test makes it difficult for researchers to determine whether or not errors are systematic or merely slips of the tongue. Moreover, researchers cannot use control groups because tailor-made tests are based on the errors that received CF. Thus a control group that received no CF cannot have a test designed for them. Without a control group, as in Siyyari (2005), there is no way to determine whether development resulted from the CF or some other factor. Finally, because all learners received both delayed and immediate CF, Varnosfadrani seems to have failed to eliminate the possibility that some of the errors that learners were tested on might have

[7] Tailor-made tests are tests consisting of items that individual learners make errors on during an activity.

received both kinds of CF. For example, if a simple past tense error was corrected in the first task with immediate CF and then another simple past tense error was provided delayed CF in the second task, then any tests on simple past tense errors might have led to right or wrong solutions based on the effects of both types of CF. In such cases, it would be impossible to claim whether immediate or delayed CF, or potentially both, led to the result.

In summary, all three of the studies reviewed here suggest, at the very least, that L2 development may result from CF regardless of the timing. However, for more confidence to be placed in the developmental effects of delayed CF and to discover more about how those effects compare with the effects of immediate CF, more research is needed

The present study has been designed to contribute to this research domain and efforts have been made to avoid the limitations of previous research by including a No CF comparison group and highly controlled CF to ensure that the timing of CF is not confounded with type of CF. Finally, the current study also addresses a gap in the timing of CF research by using a questionnaire to investigate what learners' reactions are to differentially timed CF. Even though previous CF research has found that CF in general is strongly valued by learners (Brown, 2009; Cathcart & Olsen, 1976; Chenoweth, Day, Chun, & Lupescu, 1983; Jean & Simard, 2011; Shultz, 1996, 2001), little is known about how learners react to the timing of CF.

Chapter Three

Method

In this chapter, I describe the methods used to investigate the research questions. First, the research context, participants, and design are described. Next, details are provided about the language feature, instruction, and CF administered in the study. This is followed by a description of the language measures, their scoring procedures, and reliability levels. The statistical analyses used to investigate the research questions are described next. The chapter concludes with a description of the questionnaire that was used to elicit the participants' reactions to the CF and an explanation of how it was analyzed.

3.1. Research Context and Participants

The participants in this study were recruited from three private ESL schools in a Canadian city. These for-profit schools offer ESL instruction through classes in conversational English, test preparation, grammar, pronunciation, vocabulary, and so forth. They place their students in various levels using different in-house tests, which divide the students into beginner, intermediate, and advanced groups. The schools are registered with the Canadian Association of Second Language Teachers, and as such all teachers are TESOL Canada certified. The students usually live in home stays with local families before renting their own accommodations. The majority of the students study at the schools for less than six months. They attend at least four hours of language classes every weekday.

After approaching the school administrators and receiving their consent, I visited intermediate level classes and invited students to participate in my study. Intermediate

level learners were chosen because they were considered to offer the greatest potential to demonstrate development in the target form from the CF treatments. Unfortunately, no standardized English proficiency test (e.g., TOEFL or IELTS) could be used as a means of determining their proficiency level, so it was necessary to accept students' placement into intermediate classes based upon school placement tests. Because the three schools differed with respect to placement tests and assignment of students to proficiency levels, I used additional criteria to determine whether to include/exclude students from the study. These criteria are described in more detail below.

At the end of the recruitment period, 90 students agreed to participate. They spoke a variety of first languages; the most common were Japanese, Korean, Portuguese, and Spanish, but there were also speakers of Arabic, Mandarin Chinese, and French. Participants' ages ranged from 18-30, with the majority being between 18-24 years old. There was a female to male ration of 3:1.

3.2. Research Design

This laboratory-based experimental study followed a pre- to post- to delayed post-test research design to investigate the effects of altering the timing of CF on the development of learners' knowledge of the English passive. The CF treatments were administered in one-on-one sessions with participants randomly assigned to Delayed, Immediate, or No CF (control) conditions by the True Random Number Service software (Haahr, 2006). Every participant individually followed the 3-week process outlined in Table 1. The data collection took place over 10 months and ran concurrently for different participants. Thus, on the same day that some participants were in the first week, others were in the second and third weeks.

In the first week, each participant signed an informed consent form (see Appendix A) and completed a bio-data questionnaire (see Appendix B). Then each participant completed three tests that measured knowledge of the English passive construction. These tests are described in detail below. Participants who scored above 90% or below 10% on any of the tests were eliminated, so that floor and ceiling effects were

Table 1

Overview of Research Schedule

Week 1	Week 2	Week 3
Introduction Consent form Bio-data questionnaire	Mini-lesson 10-minute English passive lesson	Delayed Post-test Test 1 (OPT A) Test 2 (AGJT A) Test 3 (ECT A)
Pre-test Test 1 (OPT A) Test 2 (AGJT A) Test 3 (ECT A)	Communicative Tasks Task 1 (Info gap) Task 2 (Story retell) Task 3 (Role play) (CF provided at different times or not at all depending upon condition)	
	Immediate Post-test Test 1 (OPT B) Test 2 (AGJT B) Test 3 (ECT B)	
	Post-instruction Questionnaire	

avoided. (Twenty-seven potential participants were excluded for scoring above 90%, and four more eliminated for scoring below 10%.[8]) In the second week, each participant was audio recorded while receiving a mini-lesson on the English passive construction (see Appendix C). Then they were asked to participate in three ten-minute communicative language tasks designed to elicit the use of the passive construction. During this time, they received one of two CF treatments (or No CF if they were in the control group) and were audio-recorded. Following the three tasks (and condition-dependent CF), the participants were given a parallel version of the three tests they had completed in the first week. After the tests, all participants completed post-instruction questionnaires on CF (see Appendix D). One week later, the participants were tested once again with the original version of the three tests they had completed in the first week.

3.3. Language Feature

This study focused on the development of learners' knowledge of the affirmative form of the English "be" passive construction in three commonly used tenses: the present simple, the simple past, and the present perfect. The passive construction was chosen because it is a feature known to take a long time for L2 learners to master (Hinkel, 2002; Izumi & Lakshmanan, 1998; Ju, 2000; Larsen-Freeman, 1995; Williams & Evans, 1998). Therefore, I expected that intermediate learners would require CF often when attempting to produce the passive. Furthermore, influenced by discussions of linguistic complexity by Hulstijn and DeGraaff (1994) and Celce-Murcia and Larsen-Freeman (1999), Spada and Tomita (2010) consider the passive to be a complex grammar feature based on the number of linguistic transformations involved in its formulation. Linguistic

[8] These eliminations did not affect the randomness of assignment because participants were randomly assigned to the instructional treatments, and these eliminations occurred at the pre-test stage.

transformations refer to the number of changes that must be made so that learners can form a structure accurately. For example, several transformations are required to form the passive version of the active sentence, “His grandfather broke the dish.” First, the object must be moved into the subject position. Then the agent (the former subject) must be prefixed with ‘by’ and placed after the verb. Next, an auxiliary verb must be added before the main verb, and that auxiliary verb must be made to agree with both the new subject and the tense of the sentence. Finally, the main verb must be conjugated into its past participle form to produce “The dish was broken by his grandfather.” The conjugation of the main verb is made even more challenging when the verb is irregular. Because several transformations are required to form the passive, it is a structure that can be observed developmentally to get a sense of how the learners progress in their interlanguage knowledge of that form.

3.4. Instruction

As shown in Table 1, in the second week, each participant individually met with me to receive instruction which consisted of a ten-minute mini-lesson on the passive and 3 ten-minute communicative tasks: an information gap task (see Appendix E), a story retelling task (see Appendix F), and a role play (see Appendix G). The mini-lesson was provided to ensure that all participants had received some explicit instruction in the passive prior to engaging in tasks where their use of the passive would be required and would be the focus of CF. The performance of the participants in the mini-lesson suggested that they all had some knowledge of the passive which they had likely learned via previous instruction. However, it was evident that there was room for improvement. Details about the passive mini-lesson and communicative tasks are provided below.

3.4.1. Mini-lesson.

Each participant was shown a picture of a dog in a fireman's arms. The participant then listened to me read a sentence while he read the same sentence written under the picture, *The fireman rescued the dog.* I then asked the participant to identify the verb, subject, and object of the sentence and to explain who or what was the receiver of the action. I explained to the participant that this sentence was in the active voice and explained that in active sentences the subject is responsible for the action of the verb. Next, I showed the participant a cropped version of the original picture in which the dog is the main focus and only the fireman's arms around the dog are visible. Then I read the following sentence below that picture which read, *The dog was rescued by the fireman.* I asked the participant to tell me the difference between the two sentences, making it clear that only the structure of the sentence had changed not the meaning. To further ensure that the participant understood this point, I provided the following explanation,

> *The first sentence is about what the fireman did. This sentence is in the active form. The second sentence is focused on the dog. The second sentence is in the passive form. In the passive, the verb is made of two parts: the verb "to be," in this case in the past tense, and the past participle of the verb, which never changes. In the passive sentence, the subject is the person or thing that received the action. It comes before the verb. The person or thing that did the action is called the "agent."*

At this point, I explained that there are two challenging points about learning the passive in English: 1) The past participles are sometimes regular and sometimes irregular, and I showed them three examples of irregular past participles from the table of irregular

past participles that the participants were given; 2) The auxiliary "be" verb must agree with the subject, and it changes with the time of the action. I explained this by drawing their attention to another table the participants were given that featured the passive in different tenses. When this explanation was finished, the participants were asked to put the two tables out of sight, so that we could begin the three communicative tasks.

3.4.2. Communicative tasks.

The three communicative tasks that followed the mini-lesson were designed to encourage the use of the aspects of the passive that were the focus of this investigation (i.e., the affirmative form of the "be" passive construction in the present simple, simple past, and present perfect tenses). These communicative tasks are the kinds of tasks regularly found in both ESL classrooms and SLA research. The order of the tasks ran from the most to the least constrained in terms of learner production. Though all three tasks have diverse content, they were introduced under the general theme of talking about change or changes.

3.4.2.1. Information gap task.

The first task was a one-way information gap task in the form of a guessing game. The task used two tables of information about inventions, discoveries, or human creations that had changed history such as paper, gravity, or the pyramids. I held one of the tables in my role as researcher and the participant held the other. The participant had to choose one item and give me hints from their table about that item until I guessed it correctly. For example, for the item "paper", the participant gave hints such as "It was invented in China" and "It was invented in 105 AD". I would then guess "paper." To ensure that the participant had to use the passive many times, my chart was missing information, so

when I did not have the information from the participant's clue, I would ask for another hint. To develop the conversational nature of the task, after I had guessed the item correctly, I would ask the participant's personal opinion on related topics such as whether paper was still useful given the advent of computers. Then I would invite the participant to give me clues about a new item. Through this one-way task, I was able to control the flow of information so that the focus of the task remained meaningful while I created contexts for the participant's use of the passive in the present simple and simple past tenses. There were 33 items in this information gap task that elicited the use of at least 10 verbs provided in a word bank that the participant could refer to. Thus, there were many chances for participants to make errors that were addressed with Immediate or Delayed CF.

3.4.2.2. Story retelling task.

The second task was a story retelling task in which a participant hears a recording of a story and then attempts to retell it. The story describes what happens as eggs change from a farm product to a food item for the breakfast table. The participant listens to the story while following along with an illustrated version of it. After hearing the story, the participant is invited to ask any comprehension questions. While the participant retells the story, he or she is permitted to refer to a word bank consisting of words from the story written in no particular order, for example, "scramble, at home, eat, customers." Though the task was focused on meaning, the nature of the story encouraged the participant to use the passive in the present simple or simple past tenses as in "the eggs are [or were] washed." Retelling this story provided many opportunities for the participant to make passive errors which were addressed with Immediate or Delayed CF.

3.4.2.3. Role play task.

The third task was a role play in which the participant played the role of a real estate agent against my role as a reluctant home buyer. The participant had to convince me to buy the house at 100 Maple Street. The participant was shown a picture of what the house looked like from the previous year when it was in terrible condition with broken windows, no driveway etc. The participant was also shown a picture of the same house after repairs, renovations, and additions. The participant was given access to a word bank with many simple present verbs and provided with a table of all the improvements that had been made (e.g., no more window problems, new driveway, and so forth). The table was accompanied by several before-and-after images illustrating many of the changes. As the reluctant home buyer, I told the participant that I was only aware of the poor condition of the house before the changes were made. I directed the participant to attempt to make me want to buy the house by explaining how it had changed. The role play was initiated with the same motivational tactic for each participant. Each participant started by having to read the following line "Mr. Quinn, you should buy 100 Maple Street." For each participant, I feigned indignation and rose to leave the room, clearly indicating that I was offended that any real estate agent would try to sell me such a terrible house as 100 Maple Street, which I had seen in the past and would definitely not buy. This tactic engaged the participant and compelled him or her to prevent me from leaving the room by saying such things as the following: "Wait! The windows have been repaired," or "Stop! The windows were repaired."

3.4.3. CF provision.

All of the CF in the communicative tasks was provided using a uniform method to

ensure that the timing of the CF, not the type of CF, was being compared. Also, the amount of CF that was provided was strictly controlled to ensure that all participants received the same amount. The following two sections provide detailed information regarding the method and amount of CF provision.

3.4.3.1. Method of CF provision.

Each provision of CF used what Choi and Li (2013) refer to as a "hybrid...corrective move" (p. 332). That is, the correction was designed to achieve the two main functions that Ellis (2006) proposes that CF serves: it contained a prompt that pushed output from the participants followed by an accurate model that provided input for the participants.

There were three reasons for using this hybrid type of CF. First, the operationalization of the CF was informed by the reactivation and reconsolidation theoretical framework discussed in Chapter 2. As explained, in this cognitive psychological framework, it is posited that when a memory is induced back into consciousness, or reactivated, that memory (or mental representation) becomes susceptible to change if it is exposed to a similar (but competing) mental representation before the mind can return it to long term memory. Arguably, CF that includes a prompt and then a model can follow the same process. The prompt causes learners to retrieve, or reactivate, their mental representation of how an utterance is supposed to be produced. Sometimes prompts alone are sufficient to assist learners in accurately repairing their original utterances. However, if the prompted mental representation is still flawed, then the accurate model that is provided by a teacher or researcher may serve as a similar (but competing) mental representation which can beneficially interfere in the reconsolidation

process of the original representation. (Alternatively, if learners accurately repair their utterance following a prompt, then the model from the teacher or researcher only serves as more positive evidence.) The second reason for using this hybrid type of CF was that it is consistent with the way that delayed CF was operationalized by some teachers in Rolin Ianziti (2010), one of the few studies in the CF literature that provides a detailed description of the way that L2 teachers operationalize delayed CF in authentic L2 pedagogy. The third reason for using this hybrid CF was that it allowed me to provide a uniform kind of CF regardless of whether that CF was given immediately or delayed until the end of a communicative task.

Because the hybrid CF in my study is novel, it is important to explain precisely how it was provided. When a participant made an error, I unobtrusively recorded it on an error mark-off sheet (see Appendix H). Provision of CF began with a prompt to elicit the incorrectly formed utterance again from the participant. Participants in the Immediate CF condition received this prompt directly after they made an error; participants in the Delayed CF condition received this prompt at the end of the communicative task. Regardless of whether the participant produced the correct or incorrect form following the prompt, I then supplied the correct model and asked participants to repeat that model. If the accurate model was repeated incorrectly, I provided it again and then asked participants to repeat it. Regardless of the accuracy on that second opportunity, no further attention was given to that error. At that juncture in the CF process, participants in the Immediate condition continued the communicative task (in which they were engaged when they made the error), while participants in the Delayed CF condition (who had already completed the communicative task in which the error was made) were given the

prompt that initiated the provision of CF for the next error they had made. Thus CF was provided within seconds to Immediate CF participants, but could be delayed from one to ten minutes for Delayed CF participants. Examples of Immediate and Delayed CF provision for each communicative task can be found in Appendix I.

3.4.3.2. Amount of CF provision.

CF was provided only on errors made on the passive construction, and the amount of CF that was provided was strictly controlled to ensure that all participants received similar amounts. If the amount of CF provided to a participant did not fall between a minimum of 6 to a maximum of 12 provisions in total over the three tasks, the data for that participant was not included in the analyses. The lower limit reflected the belief that learners could be expected to require CF at least two times in each of the three tasks. If they made less than an average of two errors per task, it was likely that they had already mastered the passive, and as such they would not have gained much from CF. On the other hand, if CF was provided even twice in each task, then participants were likely to receive some CF on the passive construction in the present simple, simple past, and present perfect tenses. The upper limit of 12 provisions of CF was set to ensure that no participants received far greater amounts of CF than others. Fortunately, only one participant had to be eliminated for not receiving enough CF, and 92% of the participants who received CF (n=55)[9] received the maximum of 12 provisions of CF. I staggered the CF so that it was nearly always evenly distributed over the three tasks, and I stopped providing CF whenever I reached 12 provisions. The error mark-off sheet allowed me to

[9] This number does not include control group participants (n=30) because they did not receive CF. However, a post-hoc analysis of their instruction treatments revealed that each control participant made at least 6 errors, indicating that they all would have received enough CF to be eligible to remain in the study had they been in either of the CF treatment conditions.

keep a record of how much CF was provided for each participant and to ensure that all participants fell within the proposed range. To ensure that the control group (who received No CF) spent the same amount of time on task, I engaged control group participants in conversation related to the contents of the tasks for any remaining time.

3.4.3.3. Pilot testing of CF provision.

In order to ensure that I would be able to provide CF in the manner and amounts that I outlined above and to ensure that the timing for all three conditions would be equal, it was essential to pilot test the instructional tasks under all three conditions (i.e., Delayed, Immediate, and No CF). After receiving ethical approval to conduct the research, I recruited 9 participants and video-taped the pilot sessions of the tasks with 3 participants in each condition. After reviewing the transcribed speech from the pilot sessions, only small changes were made (e.g., allowing participants to hear the story only once rather than twice in the story retelling task). The piloting was useful in training me to execute the tasks uniformly and to provide condition-dependent CF consistently in the correct manner.

3.5. Language Measures

To measure participants' knowledge of the passive construction, three different language measures were employed: a picture-cued OPT (see Appendix J), a timed AGJT (see Appendix K), and an ECT (see Appendix L). All tests measured knowledge of the affirmative "be" passive construction in the present simple, simple past, and present perfect tenses. These three tests were chosen for several reasons. First, they have been used in previous instructed SLA research (Spada, Jessop, Suzuki, Tomita, & Valeo, 2013; Spada et al., 2013). Second, each test measured passive knowledge in a different

modality: speaking (OPT), listening (AGJT), and writing (ECT). Furthermore the tests measure both productive (OPT and ECT) and receptive knowledge (AGJT) of the passive. All three tests were provided at three points in time: a pre-test one week prior to instruction; an immediate post-test, directly following instruction; and a delayed post-test, one week after instruction.

3.5.1. OPT.

The OPT is a picture-cued story telling test that was developed and used in Spada, Jessop, et al., 2013; Spada et al., 2013). The story is about a package that is sent to the wrong destination before arriving at the home of its intended recipient. Participants are read a preamble that explains how the package was sent and subsequently lost in the mail. They are then provided with a story board which shows the entire story in picture form with a verb in the simple present form above each picture. After their comprehension of the story is checked, the participants are asked to retell it. To help them do this, they are shown a series of separate slides with pictures that depict the story. The learner is prompted with, "What happened to the package" to describe each slide and also asked to use the verb provided on each slide. Participants usually complete the task in ten minutes.

There are two parallel versions of the OPT that contain the same items but are contextualized differently. For example, the version of the story used for the pre- and delayed post-tests involves a grandfather in Tokyo and a grandson in Vancouver, but the version used in the immediate post-test involves a mother in Mexico and a daughter in Toronto. In each version, there are nine target "passive" items and three distractor "active" items.

3.5.1.1. OPT scoring procedures and reliability.

The language that participants produced on the OPT was first transcribed and then scored. Following the scoring procedures outlined in Spada, Jessop, et al. (2013), the maximum number of points participants received per target item was three. Three points were given for a correct passive usage such as, "The package was sent to Toronto." Two points were given for partially correct attempts to produce the passive; for example, if the response contained an error on either the auxiliary verb or the formation of the past participle such as, "The package were sent to Toronto" or "The package was sended to Toronto." Participants received one point for using an accurate active construction such as, "She sent the package to Toronto."[10] Participants received no points for incorrect answers such as, "The package sent to Toronto" or when no attempt was made to correct the item.

This marking scheme was used to score the pre-, post-, and delayed post-tests for all participants. To check for consistency in scoring, a fellow graduate student separately scored the tests from 25% of the participants and the level of inter-rater reliability was good as measured by Cohen's Kappa (.90). Furthermore, I assessed the reliability of the OPT tests by using Chronbach's alpha index, which indicated a good level of reliability for the pre-test (.84) and a high level of reliability for the immediate post-test (.92) and delayed post-test (.93).

3.5.2. AGJT.

The second test is a timed AGJT developed by Spada et al. (2013). Participants

[10] While piloting the OPT, Spada, Jessop, et al. (2013) found that even native speakers occasionally supplied the active rather than the passive construction, so it was decided to provide a score for this occurrence.

are instructed to listen to a series of recorded sentences and to indicate on an answer sheet whether the item is grammatical, ungrammatical, or they do not know. Participants only hear each item once, and three seconds after they hear the item, the next one is automatically played. There are 42-items on the test, 36 of which are target items and 6 distractors. There are 18 grammatical target items and 18 ungrammatical items. Details regarding the breakdown of individual items in terms of verb and error type can be found in Appendix M. Two versions of the AGJT were used in this study: version A was used for the pre- and delayed post-tests, and version B was used for the immediate post-test. The items in the two versions are the same but the ordering is reversed, such that item 1 in version A is item 42 in version B.

3.5.2.1. AGJT scoring procedures and reliability.

The pre-, post-, and delayed post-tests of the AGJT for all participants were scored by assigning one point for accurate responses and no points for all other responses, including "Not sure." I assessed the reliability of the AGJT tests by using Chronbach's alpha index, which indicated low levels of reliability for the pre-test (.55) and the immediate post-test (.63) and an acceptable level of reliability for the delayed post-test (.71).[11]

3.5.3. ECT.

The third test is a paper and pencil ECT (Spada et al., 2013). This 24-item test has 18 incorrect target passive items and 6 incorrect distractor non-target items. Details regarding the breakdown of individual items in terms of verb and error type can be found in Appendix M. Participants are told that each item has only one error in it. They are instructed to do three things: 1) identify the incorrect part of the item, 2) correct the error,

[11] The low reliability of the AGJT is discussed in the study limitations in Chapter 5.

and 3) explain why the incorrect version is an error. Two versions of the ECT were used in this study: version A was used for the pre- and delayed post-tests and version B was used for the immediate post-test. The items in the two versions are the same but the ordering is reversed, such that item 1 in version A is item 24 in version B.

3.5.3.1. ECT scoring procedures and reliability.

Participants' responses to the ECT items were scored following the procedures outlined in Spada et al. (2013). A total of five points was possible for each item. For correctly identifying the error in the item, one point was given, such as when participants identified "making" as the error in the item "Laws are making by the government." A mark of zero was given if participants did not answer or if they identified the wrong part of the sentence, for example, if they incorrectly identified "are" as the error in the item above. Participants received two points for providing an accurate correction such as "Laws are made by the government." One point was given for partially correct attempts containing an error caused by either choosing the incorrect auxiliary verb or for incorrectly forming the past participle, for example, "Laws is made by the government" or "Laws are maked by the government." Zero points were given if participants did not attempt to correct the error or did so completely incorrectly with an answer such as "Laws making by the government." Finally, two points were given for providing a completely accurate explanation for why the error was incorrect, for example, "This is a passive sentence, so the past participle 'made' should be used." One point was given for partially correct explanations such as "This is a passive sentence, so the past participle 'maked' should be used." Zero points were given for an inaccurate, incomprehensible, or incomplete response such as "It is past tense." To check for scoring consistency, a fellow

graduate student separately scored 25% of the ECT tests. The level of inter-rater reliability was good as measured by Cohen's Kappa (.82). Furthermore, I assessed the reliability of the ECT tests by using Chronbach's alpha index, which indicated high levels of reliability for the pre-test (.89) immediate (.95) and delayed post-tests (.96).

3.6. Statistical Analyses of the Language Measures

To investigate the research questions, it was necessary to determine whether there was a change over time and whether there was a difference in that change between the three conditions. Thus, three mixed-design one-way analysis of variance (ANOVA) tests were conducted for the three language measures (i.e., OPT, AGJT, and ECT). The within-subjects factor was time (i.e., pre-, post-, and delayed post-tests). The between-subjects factor was instructional condition (i.e., Delayed, Immediate, and No CF). The dependent variables were the scores on the language measures. The critical alpha level was set at .05.

3.7. Post-instruction Questionnaire

After all immediate post-tests were administered, participants were asked to complete a questionnaire. The purpose of the questionnaire was to address the third research question: How do learners react to instruction that includes Immediate CF or Delayed CF? Thus, the questionnaire aimed at determining what the participants' reactions were to CF, including how aware they were about receiving CF and how they felt about the timing and type of CF. All of the questions had two parts: a multiple choice component that constrained their responses to a limited number of options shared by all participants, and an open-ended component that allowed participants to freely provide their individual reactions to the topic of the question (and also an opportunity to explain

they made in the multiple-choice part of the question). (See Appendix D for the questionnaire.)

3.7.1. Post-instruction questionnaire analyses.

The open-ended component of the questions were transcribed and analyzed qualitatively for emergent themes among participants' individual reactions to Immediate and Delayed CF, and also to find comments that could provide insight into their reactions. The multiple-choice component of the questions were analyzed quantitatively to identify statistical trends in the participants' multiple-choice selections.

A three-step process was used to identify statistically significant trends in the multiple-choice responses within and/or between the treatment conditions. First, I determined the percentage of times that each response was selected. For example, in the fourth question, all 30 participants in the Delayed CF treatment condition chose one of the five multiple-choice responses (i.e., A, B, C, D, E, or F)[12]. If A was chosen 15 times (out of a total of 30 choices), then I determined that A accounted for 50% of the responses. In the second step of the analysis, I compared percentages between and/or within conditions. For example, the purpose of the fourth question was to determine how participants felt about CF. One of the multiple-choice responses was "(B) I felt happy." If, for example, the percentage of participants in the Immediate CF condition that chose B was statistically significantly larger than the percentage of participants from the Delayed CF condition that chose B, this difference in percentages between the Immediate and Delayed CF participants suggested that Immediate CF participants were happier with the CF that they received than the Delayed CF participants were. The third step in the analysis was to determine whether these differences were statistically significant. To do

[12] Percentages were calculated for each of the treatment conditions separately.

so, t-tests between proportions were performed to compare the percentages and to reveal whether statistically significant differences existed between them[13]. When more than one comparison was conducted per hypothesis, Bonferroni adjustments were used to control Type 1 error. Type 2 error was controlled by the adoption of a hypothesis-wide error rate.[14]

This chapter has described the methods used to investigate the research questions. Also, the procedures were explained for data collection and analysis for the language measures and the post-instruction questionnaire. The next chapter describes the results of these analyses.

[13] All t-tests were calculated using the StatPac statistical calculator (Walonick, 2013).

[14] To illustrate the difference between a hypothesis-wide error rate and an experiment-wide error rate, consider that 5 one-way ANOVAs were performed to test separate hypotheses about the language measure scores in this research. Because this research uses a hypothesis-wide error rate, each of these tests was considered separate from the other. Each tested a separate hypothesis, so each was considered to have a .05 probability of error, or what is known as a critical alpha of .05. If this research used an experiment-wide error rate, a Bonferroni adjustment would have been made by dividing the standard critical alpha of .05 by five for each of the five statistical tests of the same type performed, and the critical alpha would have been set at .01 for each test.

There is no general consensus about whether researchers should use a hypothesis-wide or an experiment-wide error rate (Brown & Crookes, 1990; Cabin & Mitchel, 2000; Perneger, 1998, Wilson, 1962), so Cabin and Mitchell (2000) encourage researchers to explain that the choice of when to use a Bonferroni adjustment (i.e., hypothesis-wide or experiment-wide) is subjective, and that using the Bonferrroni in different ways sometimes leads to different results. Therefore, while reading my results, readers should bear in mind my decision to use Bonferroni adjustments with hypothesis-wide error rate.

Chapter Four

Results

This chapter presents the results. First, I explain the results of the tests used to check for violations of the statistical assumptions underlying the analysis of the three language measures. Then, I describe the results of the statistical analysis of the language measures. Next, I describe the qualitative and quantitative results of the analysis of the participant questionnaire.

4.1. Statistical Assumptions

Prior to conducting the mixed-design ANOVAs, assumptions for normality, homogeneity of variance, and sphericity were checked (Larson-Hall, 2010). Normality was violated by all conditions on all OPT tests, with the exception of the Immediate CF condition on the pre-test. Normality was also violated by the No CF condition on the ECT immediate- and delayed post-tests. Levene's test indicated that the assumption of homogeneity was also violated for the AGJT pre-test ($p = .04$) and immediate post-test ($p = .03$) and the OPT pre-test ($p = .01$) and immediate post-test ($p < .01$). Finally, Maulchy's test indicated that the assumption of sphericity was violated for the ECT ($p < .01$) and OPT ($p = .02$).

When the assumptions above are violated, researchers typically choose to use equivalent non-parametric analyses because those analyses do not require the same assumptions to be met. However, there is no non-parametric equivalent alternative to mixed-design ANOVA (Erlam & Loewen, 2010; Field, 2013; Larson-Hall, 2010). Therefore, following Erlam and Loewen (2010), I proceeded with the parametric analyses, recognizing that interpretations need to be made cautiously. Nonetheless, all violations of

sphericity were ameliorated by use of the Greenhouse-Geisser correction[15] (Larson-Hall, 2000). Moreover, it is widely accepted that sample sizes of 30 or more, as is the case in this study, are large enough to be robust to violations of normality that are detected in data samples because of the central limit theorem (e.g., Field, 2013; Gravetter & Walnau, 2009; Green & Salkind, 2008). As Field (2013) explains,

> ...the assumption of normality tends to get translated as "your data need to be normally distributed", even though that's not really what it means....For significant tests of models to be accurate the sampling distribution of what's being tested must be normal. Again, the central limit theorem tells us that in large samples this will be true no matter what the shape of the population. Therefore, the shape of our data shouldn't affect significance tests provided our sample is large enough. (p. 169-72)

Nonetheless, to allow for greater confidence in my conclusions, a series of non-parametric pair-wise comparisons were performed. Even though these comparisons could not serve as an equivalent method of analyses to the mixed-design ANOVA, they allowed for multiple separate comparisons to be made within groups over time and between groups at each interval of testing. The results of these non-parametric analyses supported the same conclusions as the parametric analyses, and a full description of the results is reported in Appendix N.

It is also important to report that before conducting the mixed-design ANOVAs,

[15] Violations of sphericity in repeated-measures ANOVAs increase the risk of finding an effect when one does not exist. The reason for this is that sphericity violations imply a lack of equivalence in the variances of the differences between the groups being compared, in this case, the pre-, post- and delayed post-test for the OPT and ECT, respectively. However, according to Larson-Hall (2010), when an RM ANOVA does not meet the assumption of sphericity, the Greenhouse-Geisser correction can be used to ameliorate this problem. She indicates that some authors recommend its use even when the assumption of sphericity is met, and she notes that it is more conservative than the Huynh-Feldt correction.

simple one-way ANOVAs were performed on the pre-tests of the three language measures. These ANOVAs indicated that there were no statistically significant differences between conditions prior to instruction on any of the language measures.

4.2. OPT Results

The means and standard deviations for the scores on the OPT are shown in Table 2. The results of the mixed-design ANOVA indicated that there was a statistically significant increase in scores within groups over time $F(1.85, 160.50) = 85.07$, $p < .01$, $\eta^2 = .48$. The eta squared ($\eta^2 = .48$) effect size was large.[16]

Table 2

Means and Standard Deviations for OPT Scores

Condition	Pre-test		Post-test		Delayed post-test	
	M	*SD*	*M*	*SD*	*M*	*SD*
Control (*n*=30)	10.73	7.18	16.33	8.95	17.57	8.56
Delayed (*n*=30)	8.47	7.68	21.23	5.99	18.37	8.38
Immediate (*n*=30)	6.67	4.93	17.63	8.70	16.43	9.44

Post-hoc pairwise comparisons indicated that there was a statistically significant increase in scores from pre- to immediate post-test of an average of approximately ten points (9.78, *SE*, 0.85) and from pre- to delayed post-test of an average of approximately nine points (8.83, *SE*, 0.91), but there was no statistically significant change in scores from the immediate post- to the delayed post-test.

There was also a statistically significant interaction between group and time,

[16] Eta squared is a common measure of effect size used for analysis of variance (ANOVA). Eta squared explains the percentage of variance in the dependent variable that can be attributed to the independent variable(s) (Fay & Boyd, 2010). According to Fay and Boyd (2010), "eta-squared values of .09, .14, and .22 or greater could be described in the behavioral sciences as small, medium, and large" (p. 423).

$F(3.69, 160.50) = 3.44$, $p = .01$, $\eta^2 = .04$. The eta squared ($\eta^2 = .04$) effect size was too insubstantial to even be considered small. From the data in Table 2, it appears that the group and time interaction that was detected in the mixed-design ANOVA resulted from the increase in scores from the Delayed CF condition from pre- to immediate post-test. However, follow up one-way ANOVAs conducted on the immediate- and delayed-post tests indicated no statistically significant differences between conditions.

Therefore, there was a statistically significant improvement for all conditions in their scores on the OPT from the pre-test to the immediate post-test, from the pre-test to the delayed-post test, but not from the immediate post- to the delayed post-test. Furthermore, no condition improved significantly more or less than any other condition

4.3. AGJT Results

The means and standard deviations for the AGJT scores are presented in Table 3. As was the case with the OPT, the results of the mixed-design ANOVA indicated that there was a statistically significant increase within groups over time, $F(1.96, 170.07) = 47.13$, $p < .01$, $\eta^2 = .35$. The eta squared ($\eta^2 = .35$) effect size was large. However, for

Table 3

Means and Standard Deviations for AGJT Scores

Condition	Pre-test		Post-test		Delayed post-test	
	M	*SD*	*M*	*SD*	*M*	*SD*
Control (*n*=30)	16.00	5.01	19.27	5.43	21.27	5.51
Delayed (*n*=30)	16.40	4.04	19.70	4.89	20.13	5.30
Immediate (*n*=30)	17.37	3.41	20.50	3.31	21.03	4.39

the AGJT, there was no interaction between group and time, $F(3.91, 170.07) = .86$, $p =$

.49, η^2 = .01. Unlike the OPT, the AGJT scores continued to increase with each test. Post-hoc pairwise comparisons indicated that there was a statistically significant increase from pre- to immediate post-test of an average of approximately four points (4.22, *SE*, 0.47), and from pre- to delayed post-test of an average of approximately three points (3.23, *SE*, 0.47), and also from immediate post- to delayed post-test of an average of approximately one point (0.99, *SE*, 0.42).

Thus, all participants in the instructional conditions improved their performance on the AGJT test from the pre-test to the immediate post-test, from the pre-test to the delayed-post test, and from the immediate post-test to the delayed post-test. While all improvements were statistically significant, no condition improved more or less than any other condition.

4.4. ECT Results

The means and standard deviations for the ECT scores are presented in Table 4. Once again, the one-way repeated measures ANOVA for the ECT indicated that there was a statistically significant increase in scores within groups over time $F(1.79, 155.72) = 68.26$, $p < .01$, $\eta^2 = .43$. The eta squared ($\eta^2 = .43$) effect size was large. However, there

Table 4

Means and Standard Deviations for ECT Scores

Condition	Pre-test		Post-test		Delayed post-test	
	M	*SD*	*M*	*SD*	*M*	*SD*
No CF (*n*=30)	46.37	17.22	62.03	21.33	59.53	23.48
Delayed CF (*n*=30)	44.00	16.41	54.90	18.95	58.27	18.25
Immediate CF (*n*=30)	39.43	13.53	54.27	16.48	58.03	16.90

was no interaction between group and time, $F(3.58, 155.72) = 1.44$, $p = .23$, $\eta^2 = .02$. Post-hoc pairwise comparisons indicated that there was a statistically significant increase of an average of approximately 14 points (13.80, *SE*, 1.35) from pre- to immediate post-test, and a statistically significant increase of an average of approximately 15 points (15.34, *SE*, 1.67) from pre- to delayed post-test, but there was no statistically significant increase in scores from immediate post- to delayed post-test.

Thus, all instructional conditions improved their performances on the ECT test from the pre-test to the immediate post-test, and from the pre-test to the delayed-post test, but not from the immediate post-test to the delayed post-test. However, no condition improved more or less than any other condition.

4.5. Qualitative Results of the Free Response Component of the First Question

For each of the six questions from the participant questionnaire, the results of the qualitative analysis of the free response component of each question are reported first, followed by the quantitative analysis of the trends that resulted from the multiple-choice component of each question. The first question investigated whether the timing of CF influenced participants' awareness about correction. In the free response component of the question, participants were asked if they could remember an example of CF they received. Participants' responses featured nothing that distinguished any treatment condition from any other in terms of participant awareness about CF, with the exception of the responses from the No CF condition participants. All but four of the No CF participants who made comments reported that they could not remember any CF. Unexpectedly, a few No CF participants gave examples of CF they mistakenly believed

they had received, including, “When I made a mistake about be-form,” and “house was repaired.”

Participants from the two CF conditions provided many examples of the CF that they had received. The first and most common type of response was repetition, wherein participants repeated part or all of an instance of CF they had received such as, “The eggs,” or “It was painted by Leonardo DaVinci.” The second type of response consisted of participants reporting both their mistakes and the corrections they had been provided, as in, “The window have been repaired. > The window has been repaired.” The third type of response included metalinguistic language like “Sometimes I use incorrect verb, so Paul correct my incorrect verb.” The fourth type of response indicated lack of memory such as, “I can’t remember.” Among the examples of CF reported by the participants, there were examples from all three communicative tasks, and these examples featured the passive construction in the present simple, simple past, and present perfect tenses. The participants’ examples were usually, but not always, accurate.

4.6. Quantitative Results of the Multiple-choice Component of the First Question

The multiple choice component of the first question asked, “How many times did I correct your passive mistakes?” Participants had the opportunity to choose from among the following four responses: “(A) 0-5, (B) 6-12, (C) 13-19, and (D) Over 20 times.” Table 5 shows the percentages for each of the responses for each of the conditions. To determine if there were statistically significant differences between treatment conditions in their awareness about the frequency of CF that they had received, I considered two factors: (1) differences between conditions in terms of the percentages of participants

Table 5

Percentages of Responses to the First Question of the Participants' Questionnaire

Condition	Percentages of choice of letter			
	A	B	C	D
No CF (*n*=30)	73	20	7	0
Delayed CF (*n*=30)	10	43	33	13[17]
Immediate CF (*n*=30)	10	53	13	23

from each condition that accurately chose the correct range of CF provision that they had received, and (2) differences between conditions in terms of the percentages of participants that substantially overestimated the number of times they had received CF.

As to the first factor, the participants in the No CF condition did not receive any CF, so for that condition, the accurate choice of how many times they received CF was A, a range of 0-5 times. For participants in the Delayed and Immediate CF conditions, choice B was accurate, a range of 6-12 times. As shown in Table 5, the accurate choice was made by 73% of the No CF participants, 43% of the Delayed CF participants and 53% of Immediate CF participants. As indicated in Table 6, the instructional conditions were organized in pairs, and 3 two-sample t-tests were performed between those pairs

Table 6

Two-sample t-tests Regarding Frequency of Accurate Choice

Instructional conditions	Difference in choosing accurately		
	t	*df*	*p*
No CF vs. Delayed CF	2.35	58	.44
No CF vs. Immediate CF	1.60	58	.02
Immediate CF vs. Delayed CF	0.78	58	.44

***p* < .017

[17] Rounding of percentages sometimes resulted in cumulative percentages of less than one hundred percent.

regarding accurate choice. Three tests were performed about one hypothesis, so a Bonferroni adjustment set the critical alpha level at .017. No statistically significant differences were found between conditions in the percentage of participants that accurately estimated how often they had been corrected.

As to the second factor, overestimation, the maximum number of times that any participant was corrected was 12, so choosing D (over 20 provisions of CF) was a substantial overestimation. As shown in Table 5, D was chosen by no No CF participants, 13% of Delayed CF participants, and 23% of Immediate CF participants. As indicated in Table 7, to determine whether there were any differences between the instructional

Table 7

Two-sample T-tests Regarding Substantial Overestimation

Instructional conditions	Differences in substantial overestimation		
	t	*df*	*p*
Delayed CF vs. No CF	2.04	58	.05
Immediate CF vs. No CF	2.79	58	.007**
Immediate CF vs. Delayed CF	1.01	58	.32

**$p < .017$

conditions in terms of substantial overestimation, 3 two-sample t-tests were performed. Again, a Bonferonni adjustment set the critical alpha at .017. The percentage of Immediate CF participants that substantially overestimated how often they had been corrected was larger than the percentage of No CF participants that overestimated. This was the only statistically significant difference observed between the instructional conditions.

4.7. Qualitative Results of the Free Response Component of the Second Question

The second question explored whether participants' feelings about the helpfulness of CF were influenced by the timing of the CF. In the free response section of the second

question, participants were asked to explain how helpful they thought the CF in the instructional treatment had been. As anticipated, some No CF condition participants reported that they could not remember receiving CF, commenting for example, “I don’t remember if I had corrected.” Surprisingly, however, even though they did not receive CF, some No CF participants demonstrated they believed that they had in comments such as, “I liked when you corrected because sometimes when speak I don't noticed my mistakes.” One participant’s comment suggested that some of the No CF condition participants wanted CF even if they could not remember if they had been corrected or not: “I don't remember about today, but when I made a mistake if you didn't correct my mistake, I will make mistake again.”

With the exception of those unexpected responses from the No CF participants, there was nothing in the free response comments which made any of the conditions distinguishable from the others. The most common type of response revealed how CF helped the participants notice their errors. Such comments included the following: “I noticed what is my mistakes,” and “I didn't know the sentence was correct or incorrect I can find my mistake because of corrected.” Other comments focused on the helpfulness of CF techniques, such as, “[The CF was helpful] Because I can repeat the right answer.” Still other comments revealed that participants understood that they were being corrected on their grammar such as, “I didn't know the grammar. He explain a sentence,” and “Because you help us to understand the correct form phrase and how I need to do my phrase.” Some of the grammar-related comments even included metalanguage such as, “Actually, I don't know well about passive but now I understand about passive,” “I didn't know pp + passive,” and “In Japan, I learned about pp, but I couldn't use it in my

conversation." However, there were also participants who mistakenly believed that the CF was intended to improve aspects of their speaking such as their pronunciation, "My pronunciation was too bad because I was suppose to say correct sentences sometime," or their vocabulary, "Because I could learn many vocabularies from him."

4.8. Quantitative Results of the Multiple-choice Component of the Second Question

The multiple-choice component of the second question asked, "How helpful was it when I corrected your passive mistakes?" Participants had the opportunity to choose from among the following four responses: "(A) Not helpful, (B) Somewhat helpful, (C) Very helpful, and (D) I was never corrected." Table 8 shows the percentages for each of the responses for each of the conditions. By choosing C, 33% of the No CF, 80% of the Delayed CF, and 97% of Immediate CF participants indicated that they found the CF in the treatment "very helpful." To determine whether there were any statistically significant differences between the instructional conditions in choosing C, the conditions were organized in pairs, and 3 two-sample t-tests were performed between them. The critical alpha was set at .017 due to a Bonferonni adjustment. Table 9 shows that the percentage of Immediate CF participants that chose C was larger than the percentage of No CF

Table 8

Percentages of Responses to the Second Question of the Participants' Questionnaire

Condition	Percentages of choice of letter			
	A	B	C	D
No CF (*n*=30)	7	27	33	33
Delayed CF (*n*=30)	0	20	80	0
Immediate CF (*n*=30)	0	3	97	0

Table 9

Two-sample t-tests Regarding the Helpfulness of CF

Instructional conditions	Differences in choice C		
	t	*df*	*p*
Immediate CF vs. No CF	5.20	58	< .001**
Delayed CF vs. No CF	3.67	58	< .001**
Immediate CF vs. Delayed CF	2.064	58	.04

**$p < .017$

participants that chose C. This was a statistically significant difference as was the finding that a larger percentage of Delayed CF than No CF participants chose C. There was no statistically significant difference between the percentage of Delayed and Immediate CF participants that chose C. Therefore, the percentages of participants from the CF conditions that reported that CF was very helpful were larger than the percentage of No CF participants that did so, and this difference was statistically significant.

4.9. Qualitative Results of the Free Response Component of the Third Question

The third question investigated whether participants' satisfaction about the amount of CF in the instruction was influenced by the timing of CF. The free responses reflected either participants' satisfaction or dissatisfaction with the amount of CF that they received. Among the comments that reflected satisfaction, there were no discernible patterns that revealed which of the conditions the comments originated from. However, there were three themes in the comments: (1) a positive disposition toward CF, (2) a belief that CF increases awareness of errors, and (3) a recognition that the CF was directed toward grammar errors. The majority of the comments indicated a positive disposition toward receiving CF: "I think that you corrected me according to the mistakes I had. It was very good," and "I feel that all corrections are very helpful, all thing that make me improve my English makes me feel satisfied." As was the case in the comments

in the second question, several participants focused on how CF increased participants' awareness of their errors: "I have learned something, maybe I would [not] have notice if you didn't pointed out," and "I can realize of my mistakes and for me it is easy to learn from them." Finally, there were also two comments that demonstrated awareness that the CF addressed grammar errors: "You corrected me enough sometimes I forget some grammar rules but you helped me," and "Because you help me to use a correct structure and applicate."

As is illustrated in the following examples, many No CF participants expressed dissatisfaction about the lack of CF in their instructional treatment: "If I say something wrong, please change mistakes," "I don't know if I have mistakes I want to correct," "Because I feel that I always make mistakes when I speak English, but I don't notice what and how I make mistakes," and "I'd like to listen the mistakes I make because is a way to improve my English especially at the moment of speaking." Even some participants from the CF conditions expressed a desire for more CF than they had received: "I want to know the answer of all," "more teach explanation," "It's like a bad habit. Even though you gave me a correct sentence I forgot," "Because I need help, my speaking is badly," "Because it is important more correction in this phase of learning," and "I think I made mistakes a lot."

4.10. Quantitative Results of the Multiple-choice Component of the Third Question

The multiple-choice component of the third question instructed participants to "Please choose one: (A) You did not correct me enough. I wanted more correction of my passive mistakes, (B) You corrected me enough. I was satisfied with the correction on my

passive mistakes, (C) You corrected me too much. I wanted less correction of my passive mistakes, or (D) You never corrected my passive mistakes." As is indicated in Table 10,

Table 10

Percentages of Responses to the Third Question of the Participants' Questionnaire

Condition	Percentages of choice of letter			
	A	B	C	D
No CF (n=30)	40	30	0	30
Delayed CF (n=30)	23	77	0	0
Immediate CF (n=30)	27	73	0	0

no participants from any condition chose C, and only No CF participants chose D, so to determine how satisfied the participants were with how many times they were corrected, only two factors were considered: (1) Satisfaction, represented by the percentage of participants who chose B, and (2) Dissatisfaction, represented by the percentage of participants who chose A. First, I investigated whether instructional condition affected the percentage of participants who chose B. As shown in Table 11, to determine whether

Table 11

Two-sample t-tests Regarding Instructional Conditions' Satisfaction with CF Amount

Instructional conditions	Differences in Choice B		
	t	*df*	*p*
Immediate CF vs. No CF	3.33	58	< .001**
Delayed CF vs. No CF	3.65	58	< .001**
Immediate CF vs. Delayed CF	0.36	58	.72

**$p < .017$

there was a statistically significant difference between the conditions regarding the percentage of each that chose B, the conditions were organized in pairs, and 3 two-sample t-tests were performed between the pairs of conditions. Because three tests

were required, a Bonferroni adjustment set the critical alpha level for these comparisons at .017. The percentage of participants from each of the CF conditions that chose B was larger than the percentage of No CF participants that did, and this difference was statistically significant. This indicated that the CF conditions were more satisfied than the No CF condition with the amount of CF they received. The differences in percentages between the CF groups on this choice was not statistically significant.

Turning to the second factor, dissatisfaction, as shown in Table 12, to determine whether there was a difference between the conditions regarding the percentage of each condition that chose A, the conditions were again organized in pairs, and 3 two-sample t-tests were performed. Again, the critical alpha level for these comparisons was set at .017 due to a Bonferroni adjustment. No statistically significant differences were found among the instructional conditions in terms of the percentage of participants that indicated they were dissatisfied with the amount of CF they received.

Table 12

Two-sample t-tests Regarding Instructional Conditions' Dissatisfaction in CF Amount

Instructional conditions	Differences in choice A		
	t	*df*	*p*
Immediate CF vs. No CF	1.07	58	.29
Delayed CF vs. No CF	1.42	58	.16
Immediate CF vs. Delayed CF	0.36	58	.72

**$p < .017$

4.11. Qualitative Results of the Free Response Component of the Fourth Question

The fourth question investigated whether participants' feelings about CF were influenced by the timing of CF. There were no discernible patterns in the free responses that related to individual instructional treatment. Instead, the comments revealed that participants in general had negative and positive reactions to the CF in the instructional

treatment. Negative reactions to the CF included feelings of anxiety, embarrassment, and self-disappointment. For example, one participant explained, “Because sometimes this situation is nervousing.” Anxiety also stemmed from the fear of repeating the mistakes that participants had already received CF upon: “I'm afraid of make a mistake again.” Participants reported feeling embarrassed while receiving CF, commenting, “I learned about this topic, but corrected mistake made embarrassed,” and “I felt a little bit embarrassed, but I love to someone who corrected me when I have a lot of mistakes.” Other negative reactions to CF indicated that some participants were disappointed in themselves because they required CF: “I felt I am stupid. I have same mistakes every time,” “I felt I wrong again,” and “I felt I should have known that.”

For most participants, receiving CF was a positive experience. Three themes emerged among the positive comments: general positivity, noticing, and self-improvement. Generally positive comments about CF included the following: “I like when Paul corrected the mistake,” “I felt relax. Because this way I had more confidence,” and “I'm happy to be corrected, but sometimes you could explain why I have to use and not only say the correct form.” The generally positive disposition toward CF was also revealed in comments such as, “I felt grateful” and “I felt satisfied.” The second positive theme that emerged was about noticing. One No CF participant responded to the fourth question by accurately noticing that he had not received CF and observing, “If you correct my passive mistakes, I can notice how I make mistakes, and I will try to fix my English next time.” Other comments that highlighted noticing came from participants who had received CF and who were happy that the CF had helped them notice their mistakes: “Because I didn't recognize my mistakes. So when he corrected my mistake. I

was happy," "If you didn't correct, I never notice my misunderstanding," and "I felt good and confident. It's better to know at the moment the mistakes I can remember them." Finally, participants also reacted positively to CF as a means of self-improvement: "I really want to fix my speaking, therefore, I think it's good for my skill of English," "I think is good because that makes me think that I'm improving my English, so that makes me feel happy," and "I felt comfortable. I feel good because I like learn about that."

4.12. Quantitative Results of the Multiple-choice Component of the Fourth Question

The multiple-choice component of the fourth question asked, "How did you feel while I corrected your passive mistakes?" Participants had the opportunity to choose from among the following six responses: "(A) I was anxious, (B) I was happy, (C) I was embarrassed, (D) I wanted you to finish the correction fast [henceforth referred to as impatient], (E) I felt _____ , and (F) I was never corrected." Table 13 shows the

Table 13

Percentages of Responses to the Fourth Question of the Participants' Questionnaire

Condition	Percentages of choice of letter					
	A	B	C	D	E	F
No CF (n=30)	3	40	0	0	13	43
Delayed CF (n=30)	17	40	20	0	23	0
Immediate CF (n=30)	10	58	9	3	19	0

percentages for each of the responses for each of the conditions. To gain a comprehensive understanding of how participants felt about the CF in the instructional treatment, I investigated whether there were any differences in participants' choices of emotional reactions both within and between groups. In doing so, I considered only four factors:

choices A, B, C, and D. Choice E was not statistically comparable because it was a sentence-completion answer to which participants gave different responses. Choice F was not included in any comparisons because it did not ask for an emotional reaction.

4.12.1. Within groups.

A series of one-sample t-tests between proportions were performed among the choices made by the participants in the instructional conditions to determine whether there were any statistically significant differences within each group. As can be seen in Table 13, within the No CF condition, no participant chose C or D. Thus, only one comparison was necessary, and the critical alpha level was set at .05. Choice B was selected by a larger percentage of No CF participants than choice A, $t(29) = 3.74$, $p < .01$, and this was a statistically significant difference. Thus, the percentage of No CF participants that reported that CF made them happy was larger than the percentage of them that indicated that CF made them anxious.

No Delayed CF condition participants chose D. Thus, as shown in Table 14, only three comparisons were necessary. Accordingly, the Bonferroni adjustment set the critical alpha at .017. No statistically significant differences were found among the percentages of Delayed CF participants that chose A, B, or C. Thus, there were differences between

Table 14

One-sample t-tests on Emotional Reaction to CF Choices for Delayed CF Condition

Choice	Delayed CF		
	t	*df*	*p*
A vs. B	1.75	29	.09
B vs. C	1.46	29	.15
A vs. C	0.27	29	.79

**$p < .017$

the percentage of Delayed CF participants that reported that CF made them anxious, happy, or embarrassed.

As shown in Table 15, within the Immediate CF condition, six comparisons were required, so the critical alpha was set at .0083 due to a Bonferroni adjustment. A larger percentage of Immediate CF participants chose B than A, C, or D, and this was a statistically significant finding. No differences were found among the percentage of Immediate CF participants who chose A, C, or D. Thus, the percentage of Immediate CF participants that reported that CF made them happy was larger than the percentage who reported that CF made them anxious, embarrassed, or impatient. Furthermore, there were no statistically significant differences among the percentages of Immediate CF participants who reported the latter three reactions to CF.

Table 15

One-sample t-tests on Emotional Reaction to CF Choices for Immediate CF Condition

Choice	Immediate CF		
	t	*df*	*p*
B vs. A	3.92	29	< .001***
B vs. C	4.09	29	< .001***
B vs. D	5.43	29	< .001***
A vs. C	0.13	29	.90
A vs. D	1.08	29	.29
C vs. D	0.96	29	.34

****p* < .0083

4.12.2. Between groups.

To determine whether there were any statistically significant differences between instructional conditions in terms of their emotional reactions to CF, a series of two-sample t-tests were performed. For choices A (anxiety) and C (embarrassment), three comparisons were required (i.e., No CF vs. Delayed CF, No CF vs. Immediate CF, and

Delayed CF vs. Immediate CF). Accordingly, for these comparisons the critical alpha level was set at .017 due to a Bonferonni adjustment. As shown in Table 16, no

Table 16

Two-sample t-tests Regarding Choice A (Anxiety) for Instructional Conditions

Instructional conditions	Differences in choice A		
	t	*df*	*p*
Immediate CF vs. No CF	1.10	58	.28
Delayed CF vs. No CF	1.81	58	.08
Immediate CF vs. Delayed CF	0.79	58	.43

**$p < .017$

statistically significant differences were found between the percentages of participants in the instructional conditions that indicated that CF made them anxious. As shown in Table 17, there were also no statistically significant differences found between the percentage of Immediate CF participants and the percentage of Delayed CF participants that felt

Table 17

Two-sample t-tests Regarding Choice C (Embarrassment) for Instructional Conditions

Instructional conditions	Differences in choice C		
	t	*df*	*p*
Immediate CF vs. No CF	1.68	58	.10
Delayed CF vs. No CF	2.58	58	.012**
Immediate CF vs. Delayed CF	1.21	58	.23

**$p < .017$

embarrassed by CF. This was also the case between the Immediate CF and No CF conditions. However, the percentage of participants in the Delayed CF condition that indicated they found CF embarrassing was larger than the percentage of No CF participants that indicated they found it embarrassing.

As shown in Table 13, 40 percent of Delayed CF participants and 40 percent of No CF participants chose B (happy), so no comparison was necessary between them, and

only one of the conditions needed to be compared with Immediate CF. As such, a two-sample t-test was performed between the percentages of No CF and Immediate CF participants that chose B. The critical alpha level was set at .05, and no statistically significant difference was found, $t(58) = 1.40$, $p = .17$. Thus, there were no differences in the percentages of participants from each instructional condition who indicated that CF made them happy.

As was the case for choice B above, (and as is indicated in Table 13), the same percentage of Delayed CF and No CF participants chose D (impatient), so no comparison was necessary between them, and only one condition needed to be compared with Immediate CF. Accordingly, a two-sample t-test was performed between the percentages of No CF and Immediate CF participants who chose D. The critical alpha level was set at .05, and no statistically significant difference was found, $t(58) = 0.96$, $p = .34$. Thus, there were no differences in the percentages of participants from the three instructional conditions who indicated that CF made them impatient.

4.13. Qualitatitive Results of the Free Response Component of the Fifth Question

The purpose of the fifth question was to determine what participants' preferences about the timing of CF were, and to explore whether the timing of CF provided in the instruction affected those preferences. In the free response component of the question, comments about Delayed CF such as "I can concentrate more," revealed that some participants felt that Delayed CF allowed them to focus on CF. Other participants liked Delayed CF because unlike Immediate CF, it allowed the communicative activity to proceed without interruption: "If you corrected as soon as I make a mistake, the lesson will be stopped." Some participants gave affective reasons for preferring Delayed CF that

did not interrupt classes. For example, one participant dreaded the embarrassment that Immediate CF could cause her: “If you stopped activity I was embarrassed, so result was strange.” Another participant did not wish for her errors to disrupt the class for her classmates: “In class has other student, so I trouble for them.” Finally, some participants who preferred CF to follow each task felt that such CF would be easier to remember: “I prefer to be corrected in the middle not after all activity and not after the mistake itself directly because the same I will forget the right answer if the teacher give it to me in the classical way. I will forget,” “I think I can remember more if in the end of each activity the teacher corrects me,” and “Because if it in activity I can't remember all of them.”

The majority of participants commented that they preferred CF provided immediately following an error to any kind of Delayed CF. These comments revealed affective and cognitive reasons for this preference. One affective reason for preferring CF immediately was impatience: “I want to change my mistake as soon as possible,” “I want to know right now,” and “If I'm not correct, I want to be corrected immediately.” Other comments revealed that participants’ feared embarrassing themselves by repeating the same mistake if they were not given CF immediately: “I would like to know more about my mistakes to not do them again.” A more positively-oriented affective reason for preferring Immediate CF was that some participants felt that being corrected immediately assured them that the teacher was constantly attending to them: “I prefer at the moment because I can know that I'm wrong and keep talking without be afraid of my mistakes because the teacher is going to help me,” and “As soon as I make a mistake. Teacher fixes my sentence I feel better.”

Other students provided cognitively-oriented reasons for preferring Immediate CF.

The principal cognitive factor cited was memory. Many participants maintained either that CF received immediately was the easiest type of CF to remember, or that CF that was not received immediately was too easily forgotten. The following comments came from participants who felt that immediately-provided CF was easy to remember: “If the mistake it's corrected right away it's better because I feel it's easier for me remember,” “If you correct me in the same moment I can remember the correction,” and “If someone correct my mistake as soon as I make, I can remember easily, and if after the class or little time I will forget what I said at that time.” Participants often commented that a delay in CF provision would cause them to forget their mistakes: “I can't remember my mistakes so I'd like to be corrected as soon as I make a mistake. My teacher corrects students' mistakes after each activity. But when I see the sentences on the board, I can't recognize them,” “Because I want to correct as soon as possible. I forget my mistakes if teacher was not correct soon,” and “If it's later, I may forget my mistakes.” Some participants specifically defended their choice of Immediate CF over CF provided at the end of a lesson by arguing that delaying CF until that point would make them forget their errors: “I might not remind when I made a mistake if you correct my mistakes after the lesson,” and “Because if after lesson, I can't remember the situation.” Participants also revealed that they preferred Immediate CF because the immediacy was important for helping them notice that an error had been made: “I want to be corrected as soon as possible. I don't always notice my mistake when I finished a lesson, I forget what I said,” and “It's easy to notice made mistakes because I will forget after activity or lesson.” Furthermore, some participants indicated that immediacy allowed them to better understand their errors: “If you correct me right now, I can understand why I'm mistake,

and after I will be carefully about my answer," and "I want to know mistake early because I realize mistake early if you tell me a lot of answer I confused." The final cognitively-oriented reason that many participants offered for preferring immediate correction was the belief that learning occurs during language use: "It's important to our learning been corrected in the moment when we made a mistake," "I can learn from mistake at once," "Because when I am corrected at the moment, it's easier to fixed the correct form to speak," and "Because in the time is better because this way I learn with my mistakes, if teacher corrected after the activity, I already forgot my mistakes."

4.14. Quantitative Results of the Multiple-choice Component of the Fifth Question

The multiple-choice component of the fifth question asked, "When do you like to be corrected?" Participants had the opportunity to choose from among the following four responses: "(A) In each activity, as soon as I make a mistake, (B) After each activity in a lesson, (C) After all activities in a lesson are finished, and (D) Never." As is shown in Table 18, no participants chose D, so in determining what differences there were within

Table 18

Percentages of Responses to the Fifth Question of the Participants' Questionnaire

Condition	Percentages of choice of letter			
	A	B	C	D
No CF (n=30)	83	13	3	0
Delayed CF (n=30)	70	27	3	0
Immediate CF (n=30)	87	6	6	0

and between groups, it was only necessary to consider three factors: choices A, B, and C.

4.14.1. Within groups.

A series of one-sample t-tests were performed among the percentages of choices

made by the participants in the instructional conditions to determine whether there were any differences in CF timing preferences. For the No CF and Delayed CF conditions, it was necessary to make three comparisons (i.e., A vs. B, B vs. C, and A vs. C), so the critical alpha was set at .017 due to a Bonferroni adjustment. As shown in Table 19, the

Table 19

One-sample t-tests on CF Timing Preference for No CF Condition

Choice	No CF		
	t	*df*	*p*
A vs. B	5.59	29	< .001**
B vs. C	1.41	29	.17
A vs. C	9.34	29	< .001**

**$p < .017$

percentage of No CF participants that chose A was larger than the percentage that chose B or C, and this was a statistically significant difference. However, there was no statistically significant difference between the percentages of participants that chose B and C. Thus, the percentage of No CF participants that preferred immediately-provided CF was larger than both the percentage of them that preferred CF after a task and the percentage of them that preferred CF at the end of a lesson.

As shown in Table 20, the percentage of Delayed CF participants that chose A

Table 20

One-sample t-tests on CF Timing Preference for Delayed CF Condition

Choice	Delayed CF		
	t	*df*	*p*
A vs. B	2.66	29	.013**
B vs. C	2.67	29	.012**
A vs. C	6.92	29	< .001**

**$p < .017$

was larger than the percentage that chose B or C, and the percentage of Delayed CF

participants that chose B was larger than the percentage that chose C. Both differences were statistically significant. Thus, a larger percentage of Delayed CF participants preferred CF provided immediately to any other CF timing option. Furthermore, a statistically significantly larger percentage of Delayed CF participants preferred CF provided after a task to CF provided at the end of a lesson.

As is indicated in Table 18, 6% of participants in the Immediate CF condition chose B and 6% also chose C, so no comparison was necessary between B and C, and only one of the choices needed to be compared to choice A. Thus, a one-sample t-test was performed between the percentage of Immediate CF participants that chose A and the percentage of them that chose B, and the critical alpha was set at .05. The percentage of Immediate CF participants that chose A was larger than the percentage of them that chose B (and concomitantly larger than the percentage of them that chose C), $t(29) = 8.48$, $p < .01$. This was a significant difference indicating that a larger percentage of Immediate CF participants preferred CF immediately following an error to any other CF timing option.

4.14.2. Between groups.

To determine whether there were any differences in CF timing preferences between any of the instructional conditions, a series of two-sample t-tests were performed between groups for choices A, B, and C. For choices A and B, three comparisons were required (i.e., No CF vs. Delayed CF, No CF vs. Immediate CF, and Delayed CF vs. Immediate CF). Accordingly, for these comparisons the critical alpha level was set at .017 due to a Bonferonni adjustment. As shown in Tables 21, and 22, for choices A and B, no statistically significant differences were found between the instructional conditions.

Table 21

Two-sample t-tests Regarding Choice A for Instructional Conditions

Instructional Conditions	Differences in choice A		
	t	*df*	*p*
Immediate CF vs. No CF	0.43	58	.67
Delayed CF vs. No CF	1.19	58	.24
Immediate CF vs. Delayed CF	1.60	58	.11

**$p < .017$

Table 22

Two-sample t-tests Regarding Choice B for Instructional Conditions

Instructional Conditions	Differences in choice B		
	t	*df*	*p*
Immediate CF vs No CF	0.93	58	.36
Delayed CF vs No CF	1.36	58	.18
Immediate CF vs Delayed CF	2.19	58	.03

**$p < .017$

Thus, there were no statistically significant differences among the percentages of participants from the instructional conditions regarding the preference of receiving CF immediately after an error or regarding the preference of receiving CF after each task.

As is shown in Table 18, the same percentages of Delayed and No CF participants chose C. Thus, no comparison was needed between those two conditions, and only one of the conditions needed to be compared to Immediate CF. Accordingly, a two-sample t-test with the critical alpha set at .05 was performed between Immediate CF and No CF (and concomitantly Delayed CF), and no statistically significant difference was found, $t(58) = 0.56$, $p = .58$. Thus, the participants from each instructional condition did not differ with respect to their preferences for receiving CF at the end of a lesson.

4.15. Qualitative Results of the Free Response Component of the Sixth Question

The purpose of the sixth question was to investigate whether the timing of CF

provided in the instruction affected learners' preferences about the type of CF they like to receive. In the free response component of the question, participants' comments did not feature anything that distinguished any condition from any other. Only a few participants commented that they preferred to have teachers simply provide them with the correct answer (i.e., input provision). These participants appeared to be fairly teacher-dependent, writing, "I think a teacher should give me a lot of information to improve my English skill," and "When the correct form is explained, it's easier to understand the mistake and learn the correct form." A few other participants provided more practically-minded reasons for preferring input provision, explaining that sometimes they just did not know the correct grammar, and noting, "Because sometimes I don't know the correct answer, so I think is unuseful try to guess."

Other participants indicated that they preferred to be pushed by the teacher to try to self correct (i.e., pushed output). Three themes emerged among the comments from participants that preferred pushed output: self-reliance, memory, and practice. Some participants demonstrated self-reliance, noting, "I would like to think," and "I like fix my sentences. It is more helpful for me." Other participants felt that producing or attempting to produce the language helped them to remember the correct way to use the language: "I think I can try my best to use the word well and remember them," and "Because if I try to say the right answer it hard for me to forget." Finally, other participants indicated that they believed that attempting to correct themselves provided them with the chance to and practice using accurate language: "Because I want to say and practice correct answer," "Because I like to learn myself, so I like to try more times."

Most participants preferred correction that utilized both input provision and

pushed output. Their responses came in two kinds. First there were participants who felt that both types of CF worked well separately. These participants believed that there were some situations that called for one type of CF and other situations that called for the other, or as one participant succinctly noted, “depends on the situation.” Participants who took this view often made the point that sometimes they wanted to practice, and at those times it was good to try to say the answer, but that sometimes they did not know the answer or did not know that they were making a mistake, and at those times they needed the teacher to provide the answer: “Because I can't know irregular verbs unless someone teaches me. But if I make a mistake about "be", I have to try again,” and “Actually I like to try to say the right answer, but I couldn't realize my mistakes so sometime I need to teacher's help.” Other participants had very specific reasons for wanting both types of CF in different situations as in, “In communication, I prefer that teacher gives me the right answer, in the written, I want to try to find an answer.”

The second kind of comment that favoured using two types of CF posited that both types of CF should be combined into a hybrid type of CF. Participants reported that a combination of input provision and pushed output was helpful for (1) figuring out what the accurate language was, (2) developing the ability to use that language through practice, and (3) memorizing that language. The participants who reported that it helped them figure out what the accurate language was in their own minds commented, “I want to think by myself I also want the right answer,” “I need to think the right answer again, but I also need the right answer,” and “It's important to think by myself, but I want best answer.” These participants consistently reported that they would like to think for themselves first and then get assistance from a teacher: “I want B [pushed output] first

and after A [input provision]. I can think about it more seriously,"[18] and "At first I need to try B [pushed output] and after that I want to have A [input provision] because I think I need to try to find some solution before I got correct answer from the teacher." Other participants who indicated that both input provision and pushed output were beneficial reported that it helped them develop their ability to use language accurately through practice: "Only A [input provision] > I would have never brushed my skills because I will not think by myself. Only B [pushed output] > I can't have a confidence with my answer. So both A and B are needed." Amongst these participants, some felt that it was helpful to receive the answer from the teacher first and then try themselves: "The right answer from the teachers expresses into my brain and speak again and again are two good ways improve my English," "I think teacher'd better to correct me right then. I'll try to speak again for exercise," "First teacher taught the grammar so I realize that it is incorrect. Second time I want to make correct a sentence," and "Because I want to make sure my answer and then I want to try it again." An equal number of these participants reported that trying by themselves first and then having the teachers provide them with the correct response was the best way to learn how to use the language accurately: "Try and if I can't say the teacher should say and explain the reason to use that correction," "At first I try to say the right answer because I think about my mistakes. When I can't say the right answer, the teacher gives me the right answer," "I think the first we need to try to say again the correct sentence and then teacher can correct me with the right sentence," and "I like to say the correct answer and after that the teacher tells me the mistake because next time I will know that I said correct and the teacher helped me in that mistake." Finally, some

[18] By "A" and "B," the participants are referring to the descriptions of input provision and pushed output in the multiple-choice component of the sixth question.

participants reported that the hybrid type of CF helped them memorize the accurate language: “When the teacher just gives me the right answer, and if I don't say the right answer, I don't remember that,” and “If the teacher just give me the right answer, I will forget easily, but also if just I try to say the right answer, it makes me more confuse and take a long time.” Some participants felt that memory was assisted if the teacher provided the answer first and then they attempted to use it: “If the teacher tries to explain something and after that I try to say the right answer, the explanation could be fixed on my mind.” Other participants reported that the opposite order was helpful: “It's better to memorize if I try to find the correct form alone before being corrected.”

4.16. Quantitative Results of the Multiple-choice Component of the Sixth Question

The multiple-choice component of the sixth question asked, “How do you like to be corrected?” Participants had the opportunity to choose from among the following four responses: “(A) The teacher gives me the right answer [input provision], (B) I have to try to say the right answer [pushed output], (C) Both A and B, and (D) I do not like to be corrected.” As indicated in Table 23, no participants chose D, so the statistical analyses of

Table 23

Percentages of Responses to the Sixth Question of the Participants’ Questionnaire

Condition	Percentages of choice of letter			
	A	B	C	D
No CF	7	20	73	0
Delayed CF	0	10	90	0
Immediate CF	10	13	77	0

participants’ preferences about the type of CF within and between groups were limited to choices A, B, and C.

4.16.1. Within groups.

A series of one-sample t-tests were performed among the percentages of the choices made by the participants in the instructional conditions. For all instructional conditions, it was necessary to make three comparisons (i.e., A vs. B, C vs. B, and C vs. A,), so a Bonferroni adjustment set the critical alpha at .017. As shown in Table 24,

Table 24

One-sample t-tests on CF Type Preference for No CF Condition

Choice	No CF		
	t	*df*	*p*
A vs. B	1.42	29	.17
C vs. B	3.60	29	.001**
C vs. A	5.89	29	.001**

**$p < .017$

within the No CF condition, the percentage of No CF participants that chose C (both input provision and pushed output) was larger than the percentage of No CF participants that chose A (input provision) or B (pushed output). This was a statistically significant difference but no such difference was observed between the percentage of No CF participants that chose A and the percentage of them that chose B. Thus, a larger percentage of No CF participants reported preferring both input provision and pushed output to either one on its own.

As indicated in Table 23, no Delayed CF participants chose A, so only one comparison was necessary between choices C and B, with the critical alpha set at .05. A larger percentage of Delayed CF participants chose C than B. This difference was statistically significant $t(29) = 7.30$, $p < .01$. Thus, no Delayed CF participants preferred input provision on its own, and a larger percentage of Delayed CF participants preferred both input provision and pushed output to pushed output alone.

As shown in Table 25 within the Immediate CF condition, because A, B, and C were chosen, three comparisons were required, and the critical alpha was set at .017, due to a Bonferonni adjustment. The percentage of Immediate CF participants that chose C was larger than the percentage of them that chose B or A, and this was a statistically significant finding. There was no such difference between the percentages for choices A and B. Thus, a larger percentage of Immediate CF participants reported preferring both input provision and pushed output to either type of CF on its own.

Table 25

One-sample t-tests on CF Type Preference for Immediate CF Condition

Choice	Immediate CF		
	t	*df*	*p*
A vs. B	0.24	29	.73
C vs. B	5.01	29	< .001**
C vs. A	5.56	29	< .001**

**$p < .017$

4.16.2. Between groups.

To determine whether there were any differences in the choices of CF type between the instructional conditions, a series of two-sample t-tests were performed between conditions for their choices of A, B, and C. For all those choices, three comparisons were required (i.e., No CF vs. Delayed CF, No CF vs. Immediate CF, and Delayed CF vs. Immediate CF). Accordingly, for all comparisons the critical alpha level was set at .017 due to a Bonferonni adjustment. As shown in Tables 26, 27, and 28, for choices A, B, and C, no statistically significant differences were found between the instructional conditions. Thus, there were no differences between instructional conditions regarding preferences for CF type.

Table 26

Two-sample t-tests Regarding Choice A for Instructional Conditions

Instructional Conditions	Differences in choice A		
	t	*df*	*p*
Immediate CF vs. No CF	0.42	58	.68
Delayed CF vs. No CF	1.48	58	.15
Immediate CF vs. Delayed CF	1.78	58	.08

**$p < .017$

Table 27

Two-sample t-tests Regarding Choice B for Instructional Conditions

Instructional conditions	Differences in choice B		
	t	*df*	*p*
Immediate CF vs. No CF	0.73	58	.47
Delayed CF vs. No CF	1.09	58	.28
Immediate CF vs. Delayed CF	0.36	58	.72

**$p < .017$

Table 28

Two-sample t-tests Regarding Choice C for Instructional Conditions

Instructional conditions	Differences in choice C		
	t	*df*	*p*
Immediate CF vs No CF	0.36	58	.72
Delayed CF vs No CF	1.70	58	.10
Immediate CF vs Delayed CF	1.36	58	.18

**$p < .017$

4.17. Chapter Summary

In this chapter, I reported the results of the language measures and questionnaire. In the cases of the OPT, AGJT, and ECT, there was a statistically significant improvement for participants in all instructional conditions from the pre-test to the immediate post-test and from the pre-test to the delayed post-test. It was only on the AGJT that participants in all instructional conditions improved their scores from the immediate post-test to the delayed post-test, and this was also a statistically significant

finding. No instructional condition improved more than any other condition. The questionnaire findings revealed that participants were aware of the amount of CF they had received and that regardless of instructional condition, they had positive dispositions toward CF, preferred Immediate to Delayed CF, and wanted to both be pushed toward self-correction and be provided with a model of the accurate language from their teacher. In the next chapter, I discuss how these results from the language measures and the questionnaire respond to the research questions in this study, how they relate to previous research, and their potential implications for future CF research and L2 pedagogy.

Chapter Five

Discussion

In this chapter, the research questions are discussed in relation to the results from the language measures and questionnaire. Then, the limitations of the study are described, and some potential pedagogical and theoretical implications of this research are outlined. The chapter concludes with recommendations for future research into the timing of CF.

The first research question examined whether there was a difference between the L2 grammatical development that resulted from the CF conditions and the No CF condition. The results from the language measures indicated that within groups, there was statistically significant improvement, with large effect sizes from pre- to immediate post-tests and from pre- to delayed post-tests on all language measures (and also from immediate to delayed post-tests on the AGJT), but there were no statistically significant differences between either of the CF conditions and the No CF condition.

One possible explanation for the observed increase in scores for participants in all groups may be due to a test-retest effect. That is, the learners became more skilful completing the tests because they completed the same tests (or parallel versions of the tests) more than once. However, if this had been the case, one would have expected the increases in scores to have continued over time, resulting in a statistically significant increase in scores from the immediate post-test to the delayed post-test; such an increase only occurred on the AGJT. Therefore, it is doubtful that a practice effect was responsible for all of the development that occurred, particularly in the cases of the ECT and OPT.

A more convincing explanation for the equivalent increase in scores across the conditions is that all three were provided with form-focused instruction (FFI) on the

passive-structure. The only difference between the instructional treatments was that the CF conditions received CF in addition to the FFI on the passive. Thus, the effects of the FFI on its own (i.e., without the additional CF) may have masked the effects of the CF.

Goo and Mackey (2013) have argued that the inclusion of additional instruction in CF studies makes it difficult to clearly ascribe results to the effect of CF. They suggest that this challenge of teasing out the effects that result from CF from those that result from additional FFI could be avoided by conducting studies that do not include additional FFI. In this study (as in many CF studies, e.g., Ammar & Spada, 2006; Lyster, 2004; Lyster & Izquierdo, 2009) an additional instructional treatment was required to ensure that all learners had some partial knowledge of the passive feature prior to receiving CF on their use of it. Participants who lacked that knowledge could not have gained the full benefit of the prompting component of the CF, which according to Lyster (2004) includes retrieving a previously encoded mental representation about how to use the structure from long term memory. Moreover, because the passive does not occur frequently in natural conversation, it was necessary to design FFI tasks which would elicit regular use of the passive upon which CF could be provided when necessary. To ensure that both CF and No CF conditions had similar exposure to target forms, and as such were comparable, the No CF condition had to participate in the same FFI treatments as the CF conditions (Lyster & Ranta, 2013). Li (2010) argued that such FFI-only (No CF) conditions serve well as control conditions in studies that compare CF-with-FFI conditions to FFI-only conditions because any different effects between the conditions must result from "the presence or absence of feedback" (p. 319). However, because equivalent development occurred for all conditions in this study, it was not possible to disentangle the role that the

CF played in the results from the role that the FFI played. Erlam and Loewen (2010) faced the same conundrum when the CF-with-FFI condition in their study demonstrated development over time but failed to outperform their FFI-only comparison condition. I concur with their recommendation:

> More research is needed to establish in what contexts corrective feedback leads to learning that is superior to that which results from activities in which students have the chance to focus on form while engaged in the communication of meaning but do not receive any feedback about their production. (Erlam & Loewen, 2010, p. 899)

A third possible explanation for the lack of difference between the conditions is that a longer intervention may be required before statistically significant differences emerge. Researchers regularly call for more longitudinal studies in SLA research because the processes involved in learning a second language require long periods of time (e.g., Lyster & Ranta, 2013; Ortega & Iberri-Shea, 2005). It is possible that had my study included more than a one-shot intervention, the CF conditions may have outperformed the No CF condition in time. In a longer treatment, participants would have received more than the 12 provisions of CF that were provided in this study. Perhaps more provisions of CF staggered over a longer period would have proved more conducive to learning. Future research should investigate whether or not this is the case.

The second research question explored whether there was a difference between the L2 grammatical development that resulted from the Immediate CF treatment and the Delayed CF treatment. The results from the language measures indicated there was no statistically significant difference between the two CF conditions, and that learners in

both groups followed the same pattern of improvement over time. There are several possible explanations for this result.

Some might argue that the reason for this outcome was that the two CF conditions were not different from each other from the outset because 1) they believe that the delay in the Delayed CF treatment was too short and/or 2) they do not believe that there was any delay in the Delayed CF treatment. First, it might be argued that the Delayed CF treatment may not have been delayed long enough to be different from the Immediate CF treatment, and as such that this study merely compared an Immediate CF treatment to a (slightly less) Immediate CF treatment. As explained in Chapter 2, there has been relatively little discussion about the timing of CF, and even less on the demarcation of the boundary between immediate and delayed CF. Even though Long (1977) and Chaudron (1977a) originally defined delayed correction as correction that allows learners to finish their utterances, and immediate correction as correction that interrupts them, contemporary CF research has subsumed these older notions under the label of immediate CF. In the contemporary CF literature, including this study, delayed CF refers to CF that is provided some time after a communicative task (Hunter, 2011, 2012; Rolin-Ianziti, 2006, 2010; Siyyari, 2005; Varnosfadrani, 2006). Yet, even some CF that is provided after a task might still occur shortly after an error was made, as would be the case when the errors were made in the last one or two minutes of the task. Given this possibility, how can one determine when CF ceases to be immediate? First, any construct of immediate CF should surely include CF that is immediately contiguous to an error, whether that refers to CF that interrupts learners in mid-utterance or CF that allows them to complete their utterance. However, when does immediate CF cease to be immediate?

Theoretical support can be inferred from Doughty (2001) for a demarcation point of one minute after an error is committed. As reviewed in Chapter 2, Doughty argues that if feedback (a recast, in this case) is to fit easily into working memory with the error with which it is to be compared, then that feedback should be provided within the "cognitive window of opportunity for pedagogical intervention" (p. 257), which psycholinguistic studies indicate to be "something well under one minute" (p. 257). Thus, feedback provided after this "window" has closed (i.e., after more than a minute has passed from the commission of an error) can arguably be described as delayed CF.

In future research of the timing of CF, it might be useful to consider the following conventions for distinguishing immediate and delayed CF: 1) Immediate CF refers to CF that either interrupts learners or allows learners to complete their utterances, and is provided no more than one minute after the commission of an error; 2) Delayed CF refers to CF that occurs more than a minute after the commission of an error. Even though contemporary CF research has most commonly operationalized delayed CF as CF that follows a task (i.e., end-of-task CF), delayed CF could also refer to CF that is provided at the end of a lesson (i.e., end-of-lesson CF); and finally 3) Postponed CF, in line with Long's (1977) definition of "postponed feedback" (p. 290), might best refer to CF that is provided in a subsequent lesson (e.g., Hunter, 2011, 2012)[19].

Following these suggested conventions, future researchers could investigate whether a difference in the length of delay leads to different results. Such research could investigate whether an optimal period of delay exists which is neither too short nor too

[19] Hunter refers to the CF in his research as "delayed CF", but as discussed in Chapter 2, it is debatable whether remedial instruction in a subsequent lesson which is focused on errors from a previous one should be categorized as corrective feedback. Thus, at the very least, perhaps a distinction should be made between CF that is provided within the same lesson (i.e., either immediate or delayed CF) and CF that is provided in a subsequent one (i.e., postponed CF).

long to detract from the maximum effectiveness of CF. CF that is provided immediately or provided after only a short delay may not allow learners enough time to process the CF. The immediate communicative demands of the task, or the pressure of knowing that another task will begin shortly, may rush learners and interfere with their ability to process new information. CF research has already demonstrated that some feedback which is provided immediately and is noticed by learners (at least to the extent of uptake) still sometimes fails to serve as a reliable predictor of accuracy in subsequent learner performance (Loewen & Philp, 2006; McDonough & Mackey, 2006). Moreover, Robinson (1995) argues that the noticing process that facilitates the encoding of new material in long term memory requires both detection and mental rehearsal. Perhaps CF that is provided too soon does not allow for effective mental rehearsal. Other researchers have suggested that for CF to have an optimal effect, learners may need processing time outside of the communicative pressures of communication tasks (Carroll, 2001; Hunter, 2007). Perhaps this need for processing time inspired a participant in this study to write "I can concentrate more" as an explanation for why he preferred end-of-class CF to either immediate or even end-of-task CF. On the other hand, there are reasons to believe that delaying CF too long is also problematic. In Chapter 2, I reviewed sound theoretical arguments from Doughty (2001) and Long (1977) to that end. Furthermore, it is clear that the participants in this study felt less confident in the efficacy of CF the longer that it was delayed from the commission of the original error. The conventions for distinguishing between immediate and delayed CF outlined above may provide some guidance for future research investigating whether altering the length that CF is delayed results in a difference in L2 development.

Some readers might argue that there was no delay in the Delayed CF treatment. That is, even though all Delayed CF was provided from one to ten minutes after the original commission of an error, the operationalization of CF entailed that one of the feedback moves in both the Immediate and Delayed CF conditions always immediately addressed errors, (just not the original commission of those errors). To examine this possibility, it is necessary to closely examine the way that CF was provided. In both the Immediate and Delayed CF treatments, I consistently pointed to some aspect of the communicative task (e.g., a picture in the story retelling task) and asked the participant to try to explain that aspect again. In the Immediate CF treatment, I took this step at the moment that an error occurred, whereas in the Delayed CF treatment, I waited until the task was finished before doing so. In both cases, the participant responded to the prompt by attempting a self-correction. I consistently responded to these attempts at self correction immediately by providing the participant with a model of the correct language. Therefore, in both the Immediate and Delayed CF treatments, a component of CF was provided immediately upon the commission of an error (or more precisely, the re-commission of an error during the participant's self-correction attempt). Accordingly, some might argue that the comparison made in this study was not between Immediate and Delayed CF conditions, but rather between Immediate and Delayed-immediate CF conditions.

Notwithstanding this similarity between the Immediate and Delayed CF conditions, they also differed in some obvious ways. Unlike the Immediate CF treatment, the Delayed CF treatment did not interrupt the flow of the communicative task because it was removed from the task. Moreover, the Delayed CF treatment was more intensive

because errors were corrected consecutively, whereas in the Immediate CF treatment, participants continued to engage in communication after each correction until they were interrupted by the next provision of CF.

Even with these differences between the Immediate and Delayed CF conditions some might still argue that the Delayed CF treatment was not delayed because of the immediate proximity between the error and the prompts and models given in the corrective moves. However, if near-synchronistic presence of both error and correction entails that error correction is immediate, then written CF must also be considered as immediate CF because in written CF, the error and the correction for the error are both viewed simultaneously by the learner. Nonetheless, written CF is typically thought to be an example of delayed not immediate CF. In fact, it seems that the only CF that would be labelled as truly delayed under such a strict definition of delayed CF would be CF that provided no reminder of the preceding error and was provided in a rather decontextualized manner at some juncture in time at least a minute after an error was made, and presumably after other unrelated material had been discussed. Such delayed CF would be expected to be less effective than immediate CF because more time elapsed between the commission of the error and the provision of the CF. In the case of such delayed CF, short intervals between the error and the CF might cause difficulty in terms of learners remembering what the error was, while longer intervals might lead to confusion if CF was provided out of context at a later point in the interaction. No research has operationalized delayed CF without first providing some reminder of the error, which is consistent with most teaching practices. For example, Siyarri's (2005) delayed CF consisted of explicit correction that included statements meant to remind learners of their

errors before accurate models were provided, as follows: “It is not X, but it is Y” (p. 91). Varnosfadrani (2006) also initiated his CF with reminders such as “So you said ‘he thought still gambling’” (p. 83) before allowing students to self-correct or providing them with accurate models. Hunter (2011, 2012) provided running error sheets that listed the errors that participants had made. In this study, CF purposefully prompted the participants to attempt to communicate a specific aspect of a task that they had failed to do accurately. As explained in Chapter 3, this method was adopted from the delayed CF technique Rolin-Inaziti (2010) defined as teacher-initiated student-correction. In both the delayed CF techniques in Rolin-Ianziti’s (2010) study, the teachers also used either a prompt or a model to remind learners about errors before they were corrected.

A second possible explanation for the lack of difference in the outcomes of the two CF conditions in this study is that the intervention may have been too short. Different results may have emerged after a longer treatment period that included more provisions of CF. To my knowledge, no longitudinal study has yet been conducted comparing the effects of immediate and delayed CF treatments; Varnosfadrani (2006), Sheen (2012), and Siyyari (2005) and this study all featured short treatments. Future research on the timing of CF should include longer interventions.

The third possible explanation for the lack of differences between conditions is in line with the findings of the other studies that have compared immediate and delayed CF. Put succinctly, immediate CF and delayed CF may not have differential effects on L2 development. To my knowledge, Varnosfadrani (2006), Sheen (2012), and Siyyari (2005), along with the current study are the only studies that have compared immediate to delayed CF, and the results, with one exception, have been the same: development over

time has resulted from both immediate and delayed CF treatments, but no difference between the immediate and delayed CF treatments has been found. The one exception to this pattern was Siyyari (2005) in which the immediate CF condition outperformed the delayed CF condition on an aggregate of scores despite the fact that no differences were found between the conditions for each individual language feature. However, as discussed in the Chapter 2, Siyyari's comparison was problematic because his immediate CF treatment was implicit and his delayed CF treatment was explicit. Thus, the comparison made in Siyyari's study may have been one of type rather than of timing. It will be important for future research to avoid this potentially confounding variable.

The third research question investigated the participants' reactions to Immediate and Delayed CF. Even though the different timing options of CF did not result in differential L2 development, the questionnaire responses indicated that participants did have different reactions to the two options. These reactions are discussed below in relation to the six areas explored by the questionnaire.

5.1. Noticing CF

First, the Immediate CF treatment caused participants to substantially overestimate how often they were corrected. It was not surprising that the Immediate CF participants overestimated the amount of CF they received more than the No CF participants because the No CF participants received no correction. However, one wonders why the learners from the Delayed CF condition did not overestimate how often they were corrected. This could be related to the fact that the Delayed CF was provided between communicative tasks with no communicative pressure to distract participants from noticing each time they had been corrected. Conversely, the precision of the

Immediate CF participants' memory about correction might have been marred as a result of being corrected in the midst of ongoing communicative activity; they believed that they had been corrected 66% more often than was the case. The Immediate CF participants' overestimation suggests that they did not have a clear memory of each time they were corrected, but rather had a composite memory of having received a great deal of CF. This may be related to the way in which they noticed and were aware of the CF they received.

As reviewed in Chapter 2, noticing is a central issue in the discussion about the efficacy of CF (e.g., Chaudron, 1977b; Lyster, 1998; Nicholas et al., 2001). L2 research indicates that learners' ability to notice the corrective intent of CF is important in the facilitation of L2 grammatical development. Schmidt (1990, 2001) contends that there are different levels of noticing: noticing at the level of awareness and noticing at the level of understanding. Noticing at the level of awareness involves a conscious focus of attention on something in the linguistic environment. Schmidt argues that this is the level of noticing that is minimally necessary for L2 development. Consciousness at the level of understanding is a higher level of awareness that allows for an analysis of an object of consciousness and a comparison of it with other objects of consciousness that have been noticed previously; a process that can lead to insight and metalinguistic problem solving. Schmidt argues that this higher level of noticing is facilitative but not necessary for second language acquisition.

The Immediate CF participants' substantial overestimation demonstrates that they almost certainly noticed the multiple provisions of CF at the level of awareness, but it is less clear if they noticed each provision of CF at the level of understanding. That is, they

noticed that they had been corrected multiple times, but the communicative pressures within the task may have not allowed them enough time to focus on each of the corrections to the extent that they could analyze and understand the particular points provided in each correction. As a result, their memories may have been less clear when they attempted to recall each provision of feedback in order to estimate the number of times that they had been corrected. Robinson's (1995) notion of detection and rehearsal, discussed above, may also be relevant to this discussion. Perhaps learners detected each provision of CF, but the pressures of having to complete the task constrained the amount of cognitive processing capacity that they could dedicate to mentally rehearsing each provision. As Robinson (1995) writes,

> It is possible to briefly notice and permanently or temporarily forget, and to notice and remember over time. More permanent encoding in long-term memory is a consequence of the level of activation of information in short-term memory, itself the result of rehearsal and elaboration. Thus the nature of rehearsal in short-term memory is a consequence of the processing demands of particular tasks... (p. 298-9)

One might wonder why the participants overestimated rather than underestimated the amount of CF they had received. Unfortunately, the participants provided no clues in their comments as to why this occurred. However, it seems likely that it was the intrusive nature of the Immediate CF (i.e., interruptions in the midst of communicative tasks) that led the participants to believe that it occurred with inordinate frequency. One way to explore this in more depth would be through the use of stimulated recalls.

5.2. Helpfulness of CF

Participants' reactions indicated that both CF timing options were helpful. Furthermore, and as expected, participants in both CF conditions found the provision of CF to be more helpful than participants in the No CF condition. It was surprising that some of the No CF participants reported examples of CF that they believed they had received. Perhaps they felt pressured into giving examples because the questionnaire asked for them. Both Immediate and Delayed CF participants commented that CF helped them in noticing their errors, but there was nothing in their comments that differentiated benefits of the two treatments.

5.3. Satisfaction with CF

The difference in the timing of CF also had no effect on how satisfied learners were with how much CF they had received. Once again, the only difference between the instructional conditions occurred between those who received CF and those who did not. Thus, it appears that altering the timing of CF does not affect the satisfaction that learners feel about how much CF they receive. However, some participants from both the Immediate and Delayed CF conditions commented that they wanted more CF. This finding was unexpected because CF participants received 12 provisions of CF over the three communicative tasks, or on average four corrections per ten-minute communicative task. Thus, the same participant received CF on the same grammatical feature once every two-and-a-half minutes. No teacher would likely ever provide such intensive CF in a communicative language class.

5.4. Emotional Reactions to CF

Different emotional reactions resulted from the Immediate and Delayed CF

treatments. Participants who received Delayed CF were more likely to indicate that CF was embarrassing than participants who did not receive CF. However, there was no difference in the belief that CF was embarrassing between participants who received Immediate CF and participants who did not receive CF. Furthermore, comparisons within groups indicated that the Delayed CF participants were as likely to feel happy about the CF they had received as they were to feel anxious or embarrassed. On the other hand, participants who received Immediate CF were more likely to report that CF made them happy than that it made them anxious, embarrassed, or impatient. Perhaps the isolation of the Delayed CF treatment from all other communication made learners feel more self-conscious about their errors, and/or perhaps the intensive experience of receiving four corrections consecutively led to this result. Unfortunately, the comments from the Delayed CF participants included nothing that confirmed this speculation. As was the case with the comments for all emotions, no comments from any of the instructional conditions suggested patterns unique to one condition.

Participants from all conditions reacted to CF with both negative and positive emotions. Some negative emotional reactions were expected because, as reviewed in Chapter 2, researchers who discourage teachers from using CF (e.g., Krashen, 1982; Truscott, 1999) warn that CF may undermine learners' confidence and give them a negative attitude toward language learning. Moreover, L2 teachers are sometimes reticent to provide CF because they fear negative affective reactions from learners (e.g., Basturkmen, Loewen, & Ellis, 2004; Borg, 2003).

However, when commenting about negative emotions (i.e., anxiety, embarrassment, and self-disappointment), participants often reported that these emotions

were caused by their fear of repeating an error for which they had already been corrected. Yet, without the kind of negative evidence that CF provides, learners are likely to regularly repeat their errors because they do not know that what they are saying is inaccurate, and as such they have no reason to not continue saying it. Thus, it was not surprising that even some comments about negative emotions included a positive aspect, such as in the comment from the participant who wrote, "I felt a little bit embarrassed, but I love to someone who corrected me [*sic*] when I have a lot of mistakes." There were more positive than negative comments about CF. These comments indicated that participants were more positively disposed toward CF because CF helped them notice where they were making mistakes, as reflected in the following comment: "If you didn't correct, I never notice my misunderstanding." Participants also reported feeling grateful, satisfied, and as if they could feel themselves improving through the receipt of CF.

In summary, the participants reacted more positively than negatively to CF, and there was a relationship between when CF was provided and how participants reacted to it. Immediate CF was received with happiness more than any other emotion, while Delayed CF was greeted with as much happiness as embarrassment, and anxiety. This initial finding of a relationship between the timing of CF and the emotional reaction to it should be further investigated. Once again, future research may reveal deeper insights through the use of stimulated recalls.

5.5. Preferences Regarding CF Timing

Even though the popularity of CF is well documented in the SLA literature, little is known about learners' preferences regarding the timing of CF. To my knowledge, the only such evidence comes from Bang (1999), (as cited in Loewen, Li, Fei, Thompson,

Nakatsukasa, Ahn, & Chen, 2009), who reported that learners disagreed on their preferences about when they wished to be corrected. In contrast to Bang's finding, participants in this study almost invariably indicated that they preferred that CF be provided immediately rather than delayed; the within group comparisons revealed that, regardless of their condition, larger percentages of participants preferred to receive CF immediately. There were no statistically significant differences between the three instructional conditions regarding any of the timing options.

The reasons given for preferring Immediate CF were either affectively- or cognitively- based. With respect to affect, some participants simply did not want to wait to find out what errors they had made; the impatient student who must know right away is no stranger to any experienced L2 teacher. The other affective reaction that emerged was that Immediate CF reassured some participants that the teacher was available when they were in need. This sentiment is reminiscent of related findings from the investigations of feedback in cognitive psychology. Schmidt and Bjork (1992) and Schooler and Anderson (1990) found that delaying correction so it was provided at fewer intervals in an activity was more effective in facilitating independent performance than providing constant immediate correction because the immediate feedback fostered dependence. Viewed from this perspective, Immediate CF is considered to have a potentially detrimental effect on some learners by making them too teacher dependent.

The other reasons for preferring Immediate CF were cognitively-based. Some participants commented that they believed that learning occurs in the midst of communication. This belief is not surprising given the ubiquity of the communicative approach in ESL classrooms. The most common reason for preferring Immediate CF

(cognitively-based, or otherwise) was that participants feared that any delay between their errors and CF would cause them to forget their errors. This often-articulated fear was summarized well in the following comment: “I can't remember my mistakes, so I'd like to be corrected as soon as I make a mistake. My teacher corrects students' mistakes after each activity, but when I see the sentences on the board, I can't recognize them.” Presumably, this memory problem is compounded when teachers attempt to spare learners’ feelings by providing delayed correction using generic examples of error types rather than focusing on errors actually made by individual learners.

Another interesting finding about CF timing preferences was that the within-groups comparisons indicated that only the Delayed CF participants preferred end-of-task CF to end-of-lesson CF; no participants from the Immediate CF condition chose end-of-lesson CF, and no difference in preferences for either type of delayed CF was found for the No CF participants. Perhaps the experience of having received end-of-task CF convinced some Delayed CF participants that end-of task CF was a more acceptable timing option than they might have believed otherwise. This conclusion remains speculative, however, because the comments from the Delayed CF participants did not provide any evidence to confirm it.

However, participants’ comments did reveal why a few of them preferred Delayed CF. These participants noted that Delayed CF allowed communicative activities to continue without interruption, which is often cited to be the primary benefit of delaying CF (e.g., Harmer, 2001; Long, 1977; Rolin-Ianziti, 2006). Moreover, as noted above, a few participants also reported a preference for Delayed CF because it allowed them to

focus on CF without being distracted by the communicative pressures in communicative tasks.

5.6. Preferences Regarding CF Type.

The timing of CF was unrelated to the type of CF that participants preferred. In comparisons between groups, there were no differences in type of CF preferences, and in comparisons within groups, all participants preferred receiving both input-providing (i.e., recasts) and output- pushing (i.e., prompts) CF to either type on its own. One explanation for this finding is that participants simply feel that the more CF they receive the better. However, the comments indicated that participants' preferences about CF type are more complex than this. Some participants indicated that they favoured both types of CF because they believed that some situations were served well by one type, while other situations were served well by the other. Other participants reported that they preferred when both types of CF worked together. Interestingly, half of those participants felt that input provision should be followed by output pushing, while the other half felt the order should be reversed. Nonetheless, these participants reported that this hybrid type of CF helped them think about how to correct their errors, offered them a chance to practice producing the language accurately, and best facilitated the memorization of accurate language. Overall, participants valued both types of CF, and nothing in the comments from the Delayed or Immediate CF participants indicated that they would prefer one type or the other based on differential timing.

5.7. Summary

Participants' responses to the questionnaire revealed no differences in their preferences for Immediate and Delayed CF in terms of how helpful either is perceived to

be, how satisfied learners are with either one, or how learners would like CF to be provided. Participants' responses also revealed that Delayed CF is equally as likely to lead to embarrassment or anxiety as it is to happiness, while Immediate CF is statistically significantly more likely to be reacted to more positively, and is much preferred to Delayed CF. Nonetheless, learners tended to substantially overestimate how much Immediate CF they received. Furthermore, Immediate CF may result in learners becoming overly teacher reliant.

5.8. Study Limitations

There are a number of limitations to this study that must be considered when interpreting the findings. The first limitation is the length of the intervention. A longer one might have allowed for differences to emerge between the CF and No CF treatments and/or between the two CF treatments. Unfortunately, it was impractical to add an additional day of instructional treatment for each of the 90 participants because collecting the data that was used in the study required almost one full year.

Another limitation is that there was not enough time to address the possibility of a practice effect by collecting data from a control condition that received alternative instruction and completed the same tests used in this study, what Li (2010) refers to as a "'real' control group" (p.335). In future research, I plan to address this limitation by collecting data for such a condition and comparing the results with the three conditions in my study.

A further limitation is that the low reliability scores for the AGJT pre- and immediate post-tests indicate that less confidence should be placed in the results from the AGJT than in the other tests in the study. This low reliability was unexpected because the

AGJT delayed post-test was reliable and the AGJT has been found to be reliable in a previous study (Spada et al., 2013). One possible explanation for the low reliability is that the AGJT proved so challenging to the participants that on the pre-test and immediate post-test, they primarily resorted to guessing the answers or choosing the "not sure" option.

One other limitation is also related to the language measures. The items were the same for parallel versions of the language measures that were used in this research. However, greater confidence in the reliability of the measures could have been achieved if analysis had been undertaken to establish their equivalence empirically and/or if their implementation had been counterbalanced at each testing interval and for each instruction condition in order to control for any possible minor variance associated with each version of the measures.

Finally, one might argue that the fact that the research was laboratory-based, not classroom-based thus limiting the potential generalizability of the findings to classroom contexts. As noted in Chapter 2, the effects of CF on L2 learning have resulted in greater effect sizes in laboratory-based than in classroom-based studies (see Li, 2010; Mackey & Goo, 2007). Moreover, researchers have long suggested that the noticeability of CF in laboratory-based studies, and the ability to better control for intervening variables than in the classroom, limit the transferability of laboratory-based results to the classroom (e.g., Foster, 1998; Nicholas et al., 2001; although see Gass, Mackey & Ross Feldman (2005) for a counter-argument).

I conducted a laboratory-based study because I thought doing so would allow me to control intervening variables that might prevent me from observing differential effects

from differently timed CF. Ironically, I now think that the intervening variables found in the classroom context might actually be the factors that cause differently timed CF to have different effects. It may be, for example, that Delayed CF is more noticeable in classroom contexts in which the distractions of the class cause Immediate CF to be missed by learners. Alternatively, the potential for embarrassment and anxiety from the isolated focus of Delayed CF may make it less effective than Immediate CF, which occurs during the flow of a class and does not focus as much attention on individual learners. The lack of generalizability of this study's results from the laboratory to the classroom limits how confidently pedagogical implications can be drawn from this research. Nonetheless, in the next section, I offer some tentative suggestions about CF practice in the classroom, based on the findings of this study.

5.9. Pedagogical Implications of the Study

The results of this investigation suggest that the timing of CF can be flexible and that teachers can feel confident about providing CF either immediately or after a delay without adversely affecting the development that results from CF. This is consistent with the findings from other studies that have compared the effects of immediate and delayed CF (Varnosfadrani, 2006; Sheen, 2012: Siyyari, 2005). Such flexibility is pedagogically advantageous for teachers, particularly when they need to combine instruction and CF but also need to wait until communicative activity is complete before providing CF. For example, teachers often use role plays or simulations to help learners proceduralize their ability to complete tasks. Interrupting such tasks with Immediate CF may be counterproductive to their goal. However, accuracy might be an equally vital component of the successful completion of those tasks. For example, automaticity and accuracy are

imperative for L2 students learning how to employ rhetorical devices in pubic speaking classes and for students learning English for use in areas such as air-traffic-control, where as Widdowson (2003) notes, "communicative convention requires a strict control over variation" (p. 72). If Immediate CF were the only effective means of ameliorating inaccuracy in such cases, then double the amount of instruction might be needed. That is, first, a series of simulations that included Immediate CF would be required to address errors, and then second, a series of simulations would be needed to facilitate proceduralization at the level of discourse. In contrast, a single series of simulations that employed Delayed CF could allow teachers to address both their needs for improvement in accuracy and automaticity.

The participants' responses to the questionnaire also tentatively suggest some pedagogical implications. First, if teachers wish to be in line with learners' preferences, then they must provide CF. Moreover, if they are to meet the preferences of the majority of learners, then that CF should be provided immediately. Learners' overwhelming preference for Immediate CF suggests that determining when to provide CF is a more complicated matter than is implied by Varnosfadrani's (2006) conclusion that "language teachers have no reason to neglect immediate error correction in favour of delayed error correction any more than they should neglect delayed error correction in favour of immediate correction" (p. 161). It is true that the results from the language measures in this study are in line with Varnosfadrani's claim, and furthermore that the questionnaire revealed that learners appear to perceive CF to be equally helpful irrespective of whether it is immediate or delayed. However, the questionnaire responses also indicate that when teachers consider when to provide CF, they should bear in mind that providing CF

immediately may constrain learners' level of awareness about the CF and potentially render them more teacher-dependent, while delaying CF is as likely to make learners nervous or embarrassed as it is to make them happy. Even though these pedagogical implications are derived from the views of 90 participants, it would still be premature to attempt to generalize the findings of this study beyond this particular group of participants, let alone beyond the laboratory context. More research (particularly classroom-based research) is required before any generalizations can be made.

5.10. Theoretical Implications of the Study

This thesis provides a resource for researchers seeking theoretical explanations for the effectiveness of Immediate and Delayed CF. Five theoretical frameworks were reviewed in Chapter 2. Two of those frameworks are commonly employed to explain why different types of CF might result in L2 development: immediate cognitive comparison (ICC) facilitates immediate comparisons of accurate models and learners' errors so that they can notice the differences, while prompting is argued to lead to the proceduralization of knowledge as is hypothesized in skill acquisition theory (SAT). In Chapter 2, I also introduced three complementary theoretical frameworks (i.e., preparatory attention, transfer appropriate processing, and reactivation and reconsolidation) to explain why differently timed CF might facilitate L2 development. The reason for introducing the alternatives was that it is neither generally conceded, nor plainly obvious, that the ICC and proceduralization through prompting frameworks could explain the developmental effectiveness of Delayed CF.

In the SLA literature, ICC and proceduralization through prompting are discussed in reference to CF provided at the moment in communication when learners cannot

accurately use the correct form of another language to communicate the meaning of the message that they wish to convey. Thus, it is not obvious how these frameworks could apply to CF that is provided after that moment has passed. However, one explanation for why these frameworks might explain the effectiveness of Delayed CF became apparent in my discussion of the second research question in this chapter. As discussed, some readers might conceptualize the Delayed CF in this study as being Delayed-immediate CF because one component of the researcher's CF was provided immediately upon the commission of an error (or more precisely, the re-commission of an error during a learner's self-correction attempt). In other words, despite the fact that the Delayed CF in this study was indeed delayed in time from the original commission of an error, the prompts and the models in the CF were still provided within seconds of learners becoming aware that they had made an error, and those corrective moves took place at the moment that the learner was attempting to convey meaning.

Proponents of proceduralization through prompting might not object to the idea that proceduralization through prompting could be facilitated through the kind of Delayed CF featured in this study. Ranta and Lyster (2007), for example, do not explicitly place any time limits within which prompts must occur in order for them to be effective.

It seems unlikely that proponents of ICC would concede that Delayed CF could facilitate L2 development through ICC because Delayed CF is separated in time from the discourse within which the original error was committed. Proponents of ICC appear to believe, as Hunter (2007) puts it, "that attention to form outside of the discursive context is irrelevant and therefore of no use in the acquisition process" (p. 42). Such a perspective proscribes the use of ICC as a theoretical explanation for Delayed CF, and must be

responded to in some manner by researchers wishing to support the theoretical credibility of Delayed CF, at least in terms of ICC. Even though Hunter's (2011) own results might serve as empirical evidence against such a perspective, he does not offer a detailed alternative theoretical explanation for the effectiveness of delayed CF. Instead, Hunter (2011) seems content to remain somewhat theoretically neutral on such issues, noting, "This investigation has primarily concerned itself with the pedagogical issues surrounding delayed CF, and it has therefore adopted a relatively neutral stance with regard to the more polarized theoretical positions in Second Language Acquisition" (p. 185). Siyarri (2005) is similarly silent on this important theoretical issue. Varnosfadrani (2006) argues that his form of delayed CF was an effective method of ICC because he reminded participants of errors before correcting them. Varnosfadrani (2006) further claimed that "his findings show that immediate error correction and delayed error correction are equally effective in drawing the learners' attention to discrepancies between the interlanguage and target language forms" (p. 105). Proponents of ICC could easily challenge the assertion that Varnosfadrani's (2006) operationalization of delayed CF actually facilitated what they refer to as cognitive comparison because, despite his claims to the contrary, Varnosfadrani's provisions of delayed CF were quite decontextualized. The operationalization of Delayed CF in the present study, however, cannot be so easily challenged. CF in this study was initiated by the re-creation of meaningful discourse. For each provision of CF, participants were prompted to attempt to communicate some aspect of the task they had just completed by being directed to a picture or table, and being asked to attempt to communicate something meaningful once again. Arguably, this process continued (or re-created) the communicative context in

which the original error was committed. Any errors in participants' production were then immediately met with an accurate model. Delayed CF operationalized in this manner would seem to allow for ICC to facilitate L2 development.

Whether ICC and proceduralization through prompting are acceptable as explanations for the effectiveness of Delayed CF or not, the three other frameworks proposed in Chapter 2 (i.e., preparatory attention, transfer appropriate processing, and reactivation and reconsolidation) arguably can still be used in future research to explain why L2 development might result from both Immediate and Delayed CF. The results in this study provided little evidence to oppose the use of these proposed frameworks in explaining the effectiveness of Immediate and/or Delayed CF. Thus, this thesis offers researchers a variety of possible frameworks for explaining how CF could result in L2 development regardless of whether that CF is immediate or delayed.

5.11. Conclusions and Future Directions in Research on the Timing of CF

The purpose of this study was to investigate how the timing of CF affects L2 development and to discover learners' reactions to Immediate and Delayed CF. The results indicated that statistically significant L2 development occurred over time regardless of whether CF was immediate or delayed. However, the inclusion of FFI for all instructional conditions also made it impossible to disentangle the effects of CF from the effects of FFI. The participants' reactions to the treatment of CF in the study indicate that learners wish to be corrected, and that they prefer to be corrected immediately. Furthermore, participants' reactions suggest that Immediate CF might cause constraints on the noticeability of CF and possibly result in over reliance on teachers, while Delayed CF is just as likely to make learners happy as it is to make them nervous or embarrassed.

Future research into the timing of CF should attempt to investigate these issues in more depth through the use of stimulated recalls. Furthermore, efforts should be made to disentangle the effects of Immediate and Delayed CF treatments from the effects of any other FFI that accompanies those treatments. Future studies may attempt to facilitate this outcome by providing longer interventions, which may allow for statistically significant differences to emerge. Finally, to achieve more ecologically valid findings, future studies should be conducted in classrooms with intact classes.

Research into corrective feedback is varied and represents an important stream of inquiry in the field of instructed second language acquisition. By focusing on the timing of CF, this study represents an important contribution to a domain of CF work that few researchers have empirically investigated or theorized about. Continued research will increase our knowledge and understanding about the timing of CF and further inform us about the implications for L2 teaching and learning.

References

Allwright, R. (1975). Problems in the study of the language teacher's treatment of learner error. In M. Burt & H. Dulay (Eds.), *On TESOL, 75: New directions in language learning, teachingand bilingual education* (pp. 96-109). Washington DC: TESOL.

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*(4), 543-574.

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406.

Anderson, J. (1983). *The architecture of cognition.* Cambridge: Cambridge University Press.

Anderson, J. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum.

Austin, J. (1962). *How to do things with words.* Oxford: Oxford University Press.

Basturkmen, H., Loewen, S. and Ellis, R. (2004). Teachers' stated belief about incidental focus on form and their classroom practices. *Applied Linguistics 25*, 243-272.

Bartram, M. & Walt, R. (1991). *Correction. mistake management: A positive approach for language teachers.* Hove: Language Teaching Publications.

Bitchener, J. (2012). A reflection on 'the language learning potential' of written CF. *Journal of Second Language Writing 21*, 348-363.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81-109.

Brooks, N., (1960). *Language and language learning*. 2nd edition. New York: Harcourt, Brace and World, Inc.

Brown, J. D., & Crookes, G. (1990). Research issues: The use of multiple t tests in language research. *TESOL Quarterly*, *24*(4), 770-773.

Brown, A. V. (2009). Students' and teachers' perceptions of effective foreign language teaching: A comparison of ideals. *The Modern Language Journal*, *93*(1), 46-60.

Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: when and how are the questions. *Bulletin of the Ecological Society of America*, *81*(3), 246-248.

Carroll, S.E. (2001). *Input and Evidence. The raw material of second language acquisition*. Philadelphia: John Benjamins Publishing Company.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15*(3), 357-386.

Cathcart, R. L., & Olsen, J. E. W. B. (1976). Teachers' and students' preferences for correction of classroom conversation errors. *On TESOL, 76*, 41-53.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book*. Boston: Heinle & Heinle.

Chastain, K. (1971). *The development of modern language skills: theory to practice.* Chicago, Rand McNalley.

Chaudron, C. (1977a). A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning, 27*, 29–46.

Chaudron, C. (1977b). The content and method of oral error corrrection in second language learning. A unpublished qualifiying research paper for entry into a PhD program. OISE Toronto.

Chenoweth, N. A., Day, R. R., Chun, A. E., & Lupescu, S. (1983). Attitudes and preferences of ESL students to error correction. *Studies in Second Language Acquisition*, 6, 79–87.

Choi, S. Y., & Li, S. (2012). Corrective feedback and learner uptake in a child ESOL classroom. *RELC Journal, 43*(3), 331-351.

Cohen, A. D. (1975). Error correction and the training of language teachers. *The Modern Language Journal, 59*(8), 414-422.

Corder, S. P. 1967. The significance of learners' errors. *International Review of Applied Linguistics 5*, 161-9.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*(2), 195-221.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practising second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). New York: Cambridge University Press.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-151). New York: Cambridge University

Press.

DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology.* New York: Cambridge University Press.

Doughty, C. (1994a). Finetuning of feedback by competent speakers to language learners. In J. Alatis (Ed.), *GURT 1993* (pp. 96–108). Washington, DC: Georgetown University Press.

Doughty, C. (1994b, March). *Finetuning of L2 input: A crosslinguistic study.* Paper presented at the annual meeting of the American Association of Applied Linguistics, Baltimore, MD.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206-257). New York: Cambridge University Press.

Doughty, C., & Pica, T. (1986). "Information gap" tasks: Do they facilitate second language acquisition?. *TESOL quarterly*, *20*(2), 305-325.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). Cambridge: Cambridge University Press.

Edge, J. (1989). *Mistakes and correction.* New York: Longman.

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness 13*, 243–264.

Ellis, N. C. (2005). At the interface: Dynamic interactions explicit and implicit language knowledge. *Studies in Second Language Acquisition 27,* 305-52.

Ellis, R. (1985). Teacher-pupil interaction in second language development. *Input in second language acquisition*, 69-85.

Ellis, R. (2001). Introduction: Investigating form-focused instruction. *Language Learning*, *51*(Supplement 1), 1-46.

Ellis, R. (2006). Researching the effects of form-focussed instruction on L2 acquisition. *AILA Review*, *19*(1), 18-41.

Ellis, R. (2007a). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 339-360). Oxford: Oxford University Press.

Ellis, R. (2007b). *Cognitive, social and psychological dimensions of corrective feedback* Paper given at conference on Sociocognitive Aspects of Second Language Learning and Teaching, University of Auckland.

Ellis, R. (2012). *Language teaching research and language pedagogy*. Malden, MA: Blackwell.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning, 51*(2), 281-318.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*(2), 339-368.

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, *28*(4), 575.

Erlam, R., & Loewen, S. (2010). Implicit and explicit recasts in L2 oral French interaction. *Canadian Modern Language Review*, *66*(6), 877-905.

Fanselow, J. (1977). The treatment of error in oral work. *Foreign Language Annals, 10*(4), 583-593.

Fay, K., & Boyd, M. (2010). Eta-Squared. In Neil J. Salkind (Ed.), *Encyclopedia of Research Design.* (pp. 423-426). Thousand Oaks, CA: SAGE Publications, Inc.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles, California: Sage.

Forcato, C., Burgos, V. L., Argibay, P. F., Molina, V. A., Pedreira, M. E. & Maldonado, H. (2007). Reconsolidation of declarative memory in humans. *Learning and Memory*, *14*, 295–303.

Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied linguistics*, *19*(1), 1-23.

Gass, S. & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S., Mackey, A., & Ross−Feldman, L. (2005). Task−Based interactions in classroom and laboratory settings. *Language Learning*, *55*(4), 575-611.

Gass, S. M., & Varonis, E. (1985). Variation in native speaker speech modification to non-native speakers. *Studies in Second Language Acquisition*, *7*(1), 37-58.

Goo, J., & Mackey, A. (2013). The case against the case against recasts.*Studies in Second Language Acquisition*, 1-39.

Gravetter, F. J., & Wallnau, F.B. (2009). S*tatistics for the behavioral sciences*. Belmont, CA: Cengage Learning.

Green, S. B., & Salkind, N. J. (2008). *Using SPSS for Windows and Macintosh: Analyzing and understanding data, 5th ed*. Upper Saddle, NJ: Pearson.

Haahr, M. (2006). Random.org: True random number service Random generator [Software]. Retrieved from http://www.random.org.

Harmer, J. (2001). *The practice of English language teaching.* Essex: Pearson Education Limited.

Harris, R. (1998). Making grammar instruction relevant through student-run conversations. Paper presentation at the TESOL Convention, Seattle, Washington.

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.

Hendrickson, J. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *The Modern Language Journal*, *62*(8), 387–98.

Hinkel, E. (2002). Why English passive is difficult to teach (and learn). In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 233-259). Mahwah, NJ: Lawrence Erlbaum.

Holley, F. M., & King, J. K. (1971). Imitation and correction in foreign language learning. *The Modern Language Journal, 55*(8), 494-498.

Hondo, J. (2009). *Constructing knowledge in SLA: The impact of Timing in Form-Focused Intervention.* Unpublished Ph.D. Dissertation, Lancaster University, UK.

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion. *Memory, 17*(5), 502-510.

Hunter, J.D., (2007). *The question of error.* Unpublished Ph.D. Dissertation (Module 1), University of Birmingham, UK.

Hunter, J.D., (2011). *A multi-method investigation of the effectiveness and utility of delayed corrective feedback in second language oral production*. Unpublished Ph.D. Dissertation (Module 3), University of Birmingham, UK.

Hunter, J. (2012). ‘‘Small talk’: Developing fluency, accuracy, and complexity in speaking,’ *ELT Journal*, 66 (1), 30-41.

Hupbach A., Gomez, R., Hardt O., Nadel L. (2007). Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learning & Memory, 14*(1), 47–53.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*, 97–112.

Izumi, S., & Lakshmanan, U. (1998). Learnability, negative evidence and the L2 acquisition of the English passive. *Second Language Research, 14*(1), 62-101.

Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: Necessary, but Boring?. *Foreign Language Annals*, *44*(3), 467-494.

Ju, M.K. (2000). Overpassivization errors by second language learners, *Studies in Second Language Acquisition, 22*, 85–111.

Krashen, S. D. (1982). *Principles and practice in second language learning.* NY: Pergamon.

Krashen, S. D. (1985). *The input hypothesis.* London: Longman.

Krashen, S. D., Seliger, H. W. (1975). The essential contributions of formal instruction in adult second language learning. *TESOL Quarterly, 9*, 173-183.

LaBerge, D., Auclair, L.& Seiroff, E. (2000). Preparatory attention: Experiment and theory. *Conciousness and Cognition, 9*(3), 396-434

LaPierre, D. (1994). *Language output in a cooperative learning setting: Determining its effects on second language learning.* Unpublished MA thesis, University of Toronto.

Larsen-Freeman, D. (1995). On the teaching and learning of grammar: Challenging the myths. In F. Eckman, D. Highland, P. Lee, J. Mileham & R. Rukowski Weber (Eds.), *Second language acquisition theory and pedagogy*. Mahwah, NJ: Lawrence Erlbaum.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Taylor & Francis.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*(2), 309-365.

Lightbown, P. M. (1998). The importance of timing in focus on form. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 177-196). Cambridge, UK: Cambridge University Press.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom

second language acquisition. In Z. H. Han (Ed.), *Understanding second language process* (pp. 27-44). Clevedon, UK: Multilingual Matters.

Lin, Y. H., & Hedgcock, J. (1996). Negative feedback incorporation among high-proficiency and low-proficiency Chinese-speaking learners of Spanish. *Language Learning, 46*, 567-611.

Loewen, S. (2004). Uptake in incidental focus on form in meaning-focused ESL lessons. *Language Learning, 54*(1), 153-188.

Loewen, S., Li, S., FEI, F., Thompson, A., Nakatsukasa, K., Ahn, S., & Chen, X. (2009). Second language learners' beliefs about grammar instruction and error correction. *The Modern Language Journal, 93*, 91-104.

Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *Modern Language Journal, 90*(4), 536-556.

Long, M. H. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition*, *5*(2), 177-193.

Long, M. H. (1977). Teacher feedback on learner error: Mapping cognitions In H. D. Brown, C. A. Yorio, & R. H. Crymes (Eds.), *On TESOL* '77 (pp. 278-293). Washington: TESOL. 278-294

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective*. (39-52). Amsterdam: John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T, Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). New York: Academic Press.

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Erlbaum.

Lyster, R. (1994). La ne´gociation de la forme: Strate´gie analytique en classe d'immersion. *Canadian Modern Language Review, 50*, 447–465.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition, 20*(1), 51-81.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*(3), 399-432.

Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, *59*(2), 453-498.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition, 28*(2), 269-300.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19*(1), 37-66.

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition, 35,* 167-184.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition, 32*(2), 265-302.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(01), 1-40.

Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition, 21*(4), 557-587.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive implicit negative feedback? *Studies in Second Language Acquisition, 22*(4), 471-479.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford: Oxford University Press.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal, 82*(3), 338-356.

Mackworth, N. H. (1950). Researches on the measurement of human performance. *Medical Council Special Report 168*.

Morris, F. A. (2002). Negotiation moves and recasts in relation to error types and learner repair in the foreign language classroom. *Foreign Language Annals, 35*(4), 395-404.

Morris, D. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533.

Nader, K., & Einarsson, E. O. (2010). Memory reconsolidation: an update. *Annals of the New York Academy of Sciences. 1191*, 27–41.

Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning, 57*(4), 511-548.

Nicholas, H., Lightbown, P. M., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*(4), 719-758.

O'Relly, L. V., Flaitz, J., & Kromrey, J. (2001). Two modes of correcting communicative tasks: Recent findings. *Foreign Language Annals, 34*(3), 246-252.

Ohta, A. (2000). Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese classroom. In J. K. Hall & L.Verplaeste Eds. , *The construction of second and foreign language learning through classroom interaction* (pp. 47–71). Mahwah, NJ: Erlbaum.

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition, 17*(4), 459-481.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics, 25*, 26-45.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*(4), 573-595.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ: British Medical Journal, 316*(7139), 1236.

Pica, T. (1988). Interlanguage adjustments as an outcome of NS-NNS negotiated interaction. *Language Learning, 38*(1), 45-73.

Pica, T., Holliday, L., Lewis, N., & Morgenthaler. L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11*(1), 63–90.

Pica, T., Lincoln-Porter, F., Paninos, D., & Linnell, J. (1996). Language learners' interaction: How does it address the input, output, and feedback needs of L2 learners? *TESOL Quarterly, 30*(1), 59-84.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The awareness‐practice‐feedback sequence. In R. M. DeKeyser (Ed.) *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141‐160). Cambridge, UK: Cambridge University Press.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283-331.

Roediger, H. L. (2013). Applying cognitive psychology to education translational Educational Science. *Psychological Science in the Public Interest, 14*(1), 1-3.

Rolin-Ianziti, J. (2006). Teacher corrective practices in the foreign language classroom: The effect of timing. In C. Bailey & K. Barnett (Eds.), *Proceedings of the Social Change in the 21st Century 2006 Conference*, Queensland University of Technology: Centre for Social Change Research.

Rolin-Ianziti, J. (2010). The organization of delayed second language correction. *Language Teaching Research, 14*(2), 183-206.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar: A meta-analysis of the research. In J.M. Norris & L. Ortega (Eds), *Synthesizing research on language learning and teaching* (pp. 133–164). Amsterdam: John Benjamins.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–82.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial in the role of attention and awareness in language learning. In R. Schmidt (Ed.) *Attention and awareness in language foriegn language learning*, (pp. 1-64). Honolulu, HI: University of Hawaii Press

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Schooler, J., & Anderson, J. R. (1990). *The disruptive potential of immediate feedback.* Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, 702-708. Cambridge: MA

Schultz, R. (1996). Focus on form in the foreign language classroom: Students' and teachers' views on error correction and the role of grammar. *Foreign Language Annals*, 29, 343–364.

Schulz, R. (2001). 'Cultural differences in students and teachers perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia,' *Modern Language Journal*, 85 (2), 244-258.

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*(2), 147-163.

Segalowitz, N., & Lightbown, P. M. (1999). Psycholinguistic approaches to SLA. *The Annual Review of Applied Linguistics, 19,* 23–43.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8*(3), 263-300.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10*(4), 361-392.

Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 301-322). Oxford: Oxford University Press.

Sheen, Y. (2008). Recasts, language anxiety, modified output, and L2 learning. *Language Learning, 58*(4), 835-874.

Sheen, Y. (2010a). Differential effects of oral and written corrective feedback in the ESL classroom *Studies in Second Language Acquisition, 32,* 203-234.

Sheen, Y. (2010b). The role of oral and written corrective feedback in SLA *Studies in Second Language Acquisition, 32,* 169-179.

Sheen , Y. , Wright , D. , & Moldawa , A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System, 37,* 353 – 371.

Sheen, Y. (2012). The timing of corrective feedback and L2 learning. Paper presentation at the Second Language Research Forum, Carnegie Mellon University, U.S.A.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Siyyari, M. (2005). *A comparative study of the effect of implicit and delayed, explicit focus on form on Iranian EFL learners' accuracy of oral production.* Unpublished MA thesis, Iran University of Science and Technology.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 30*, 73-87.

Spada, N. (2006). Communicative language teaching: Current status and future prospects. In J. Cummins & C. Davis (Eds.), *The international handbook of English language teaching*. Norwell, MA: Springer.

Spada, N., & Fröhlich, M. (1995). *COLT: Communicative orientation of language teaching observation scheme: Coding conventions and applications.* Sydney,

Australia: National Centre for English Language Teaching and Research.

Spada, N., Jessop, L., Suzuki, W., Tomita, Y. & Valeo. (2013). Isolated and integrated form-focused instruction: Effects on different types of L2 knowledge. *Unpublished manuscript.*

Spada, N., & Lightbrown, P. M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15*(2), 205-224.

Spada, N., & Lightbown, P.M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*(2), 181-207.

Spada, N., Shiu, Li –Ju, & Tomita, Y. (2013). A validation of elicited imitation as a measure of implicit knowledge: Comparisons with other L2 measures. *Unpublished manuscript.*

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning 60*(2), 263-308.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

Tomasello, M., & Herron, C. (1988). Down the garden path: Inducing and correcting overgeneralization errors in the foreign language classroom. *Applied Psycholinguistics, 9*(3), 237-246.

Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: The garden path technique. *Studies in Second Language Acquisition, 11*(4), 385-395.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning 46*, 327-369.

Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review, 55*(4), 437-456.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1-41.

Varnosfadrani, A. D. 2006. *A comparison of the effects of implicit/ explicit and immediate/delayed corrective feedback on learners' performance in tail*or-*made tests*. Unpublished PhD thesis, University of Auckland.

Walker, M. P., Brakefield, T., Hobson, J. A. & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425*, 616–620.

Walonick, D. S. (2013). StatPac Statistics Calculator [Software]. Retrieved from http://www.statpac.com/statistics-calculator/percents.htm.

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second-language competence. *Applied linguistics,8*(2), 95-110.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7*(2), 133-161.

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12*(4), 416-432.

Widdowson, H.G. (2003). *Defining issues in English language teaching*. Oxford University Press, Oxford, England.

Williams, J., & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139–155). Cambridge: Cambridge University Press.

Williams, J. (2012). The potential role (s) of writing in second language development. *Journal of Second Language Writing 21*, 321-331.

Wilson, W. (1962). A note on the inconsistency inherent in the necessity to perform multiple comparisons. *Psychological Bulletin*, *59*, 296–300.

Wolff, P. & Ventura, T. (2009). When Russians learn English: How the semantics of causation may change. *Bilingualism: Language and Cognition 12*(2), 153–176.

Yang, Y., & Lyster, R. (2010). Effects of oral production practice and feedback on EFL learners' acquisition of regular and irregular past-tense forms. *Studies in Second Language Acquisition, 32*(2), 235-263.

Yoshida, R. (2008). Teachers' choice and learners' preference of corrective feedback types. *Language Awareness, 17*(1), 78-93.

Appendix A
Informed Consent Form

UNIVERSITY OF TORONTO
OISE | ONTARIO INSTITUTE
FOR STUDIES IN EDUCATION

Dear Student, XXXX YY, 2011

I am a PhD student at the University of Toronto researching the teaching and learning of English as a second language. I am writing to ask if you would be willing to participate in my research. I am offering a forty-five-minute workshop in which you will take 3 English language tests. The tests include grammar, listening, and speaking and will be done with me individually in a room in a University of Toronto library. For the speaking tests, you will be audio-recorded. I will provide you with a summary of the research if you are interested.

At the end of the workshop, I may invite you to meet with me once again for 2 further workshops with language instruction and testing in the following 2 weeks. If you do very poorly or very well in the first workshop, then I will not invite you back because I am only looking for students who do moderately well. If I do not invite you, I will give you a five-dollar-pre-paid-card for Tim Horton's Coffee Shops. If I do invite you, I will give you a five-dollar-pre-paid-card for Tim Horton's Coffee Shops after each time we meet. If you are invited to return, you will be audio-recorded during instruction and asked to meet with me again for a second seventy-minute workshop 1 week later, and a third forty-minute workshop the week after that.

If you are interested in participating in this research project, I will ask you to sign the consent form on the next page allowing me to use the information from your recordings and test scores in my PhD research and in journal articles or conference presentations about the research. Also, I will ask you to fill in a biographical questionnaire about your gender, age, nationality, first language, and experience studying English. No identifying information about you will be seen by anyone other than my supervising professor, Dr. Nina Spada, and I. Your name will be substituted for a code number. All information and recordings will be password protected or locked in my home office and destroyed no later than seven years from now.

This research has no connection to your language school. Finally, you are free to choose to not participate at any time. If you do so, your information will be erased or shredded and not used in my research, but you will still receive a five-dollar-pre-paid-card for Tim Horton's Coffee Shops for each time that we meet.

If you have any questions or concerns, please contact me: Paul Quinn, PhD. Candidate, OISE/University of Toronto, Email: paul.quinn@utoronto.ca Phone: 647 201 0024; or contact my supervising professor, Dr. Nina Spada, at nina.spada@utoronto.ca ; or contact the Research Ethics Board at the University of Toronto **ethics.review@utoronto.ca**

UNIVERSITY OF TORONTO
OISE | ONTARIO INSTITUTE
FOR STUDIES IN EDUCATION

I have read Paul Quinn's letter describing the goals of the research project, and I understand that my participation in this research will involve some or all of the following activities and conditions.

Activities

- participate in **1** of the options involving workshop(s) of English language teaching and testing:
  1. Option 1: 1 forty-five-minute workshop only;

     or

  2. Option 2: 1 forty-five-minute workshop, a second seventy-minute workshop 1 week later, and a third forty-minute workshop the week after that
- be audio-recorded during the speaking tests and instruction

Conditions

- a five-dollar-pre-paid-card for Tim Horton's Coffee Shops if I am only invited to participate in option 1, and I attend the forty-five minute workshop
- a five-dollar-pre-paid-card for Tim Horton's Coffee Shops each time I attend a research workshop if I am invited to participate in option 2
- All information about my involvement in the research will be kept strictly confidential.
- No identifying information will be made public in any publications or research reports based on this research.
- Participation, non-participation, or early withdrawal in this research and the results of my tests will not be known or used in anyway in the evaluation of my performance in any class that I am enrolled in
- I am free to withdraw from the research at any time.

Yes, I agree to participate in the research _____(Please check)

Name: _____

Email: _____

Signature: _____

Date: _____

Appendix B

Bio-data Questionnaire

Name:_____

Section 1.

Please circle one answer for the following questions:

1, I am male, female.

2. I am 18-24, 25-29, over 30 years old.

3. I am level _____ at _____.

Section 2.

Please complete the following sentences:

1. My native language is (first language)_____.

2. I am from (country name)_____.

3. I have studied English for (number of months or years) _____.

Appendix C
Passive Mini-lesson

Participants are shown the following sentence:
"The fireman rescued the dog."

Participants are asked,
"What is the verb? What is the subject of the sentence? What is the object? Who/What is the receiver of the action?"

Participants are told that this sentence is known as active and explain,
"The subject is responsible for the action of the verb."

Participants are shown the following sentence below the first one,
"*The dog was rescued by the fireman."*

Participants are asked,
"What is the difference? What is each sentence about? Who does the action? Who/What receives the action?"

Responses are elicited and then following explanation is provided,
The first sentence is about what the fireman did. This sentence is in the active form. The second sentence is focused on the dog. The second sentence is in the passive form. In the passive, the verb is made of two parts: the verb "to be", in this case in the past tense, and the past participle of the verb, which never changes. In the passive sentence, the subject is the person or thing that received the action. It comes before the verb. The person or thing that did the action is called the "agent".

A table with many past participles of irregular verbs is provided.

Base	Past Participle
be	been
become	become
begin	begun

Finally, participants are given a table with some examples and told how the passive is constructed for the present simple, simple past, and present perfect tenses.

Simple Present: verb 'to be' + past participle (pp) (+by)
The museum **is visited** by many tourists every day.
Simple Past: was/were + PP (+by)
The lady **was bitten** by her cat.
Present Perfect: has/have + been + PP (+by)
The door **has been repaired** several times this year.

Appendix D
Post -instruction Questionnaire

Name:_____ **Date:**_____

The questions below are about the three activities (information gap, story-retelling, and role play) we did together in the lesson today.

1. How many times did I correct your **passive mistakes**?
   A. (0-5)
   B. (6-12)
   C. (13-19)
   D. (Over 20 times)

Please write one example of a passive mistake that I corrected for you.

_____
_____

2. How helpful was it when I corrected your **passive mistakes**?
   A. Not helpful
   B. Somewhat helpful
   C. Very helpful
   D. I was never corrected.

Please explain.

_____
_____

3. Please choose one:
   A. You did not correct me enough. I wanted more correction of my **passive mistakes**.
   B. You corrected me enough. I was satisfied with the correction on my **passive mistakes**.
   C. You corrected me too much. I wanted less correction of my **passive mistakes**.
   D. You never corrected my **passive mistakes**.

Please explain your choice.

_____
_____

4. How did you feel while I corrected your **passive mistakes**?
   A. I was anxious.
   B. I was happy.
   C. I was embarrassed.
   D. I wanted you to finish the correction fast.
   E. I felt
   F. I was never corrected.

Please explain your choice.

_____
_____

5. When do you like to be corrected?
   A. In each activity, as soon as I make a mistake.
   B. After each activity in a lesson.
   C. After all activities in a lesson are finished.
   D. Never

Please explain.

_____

_____

6. How do you like to be corrected?
   A. The teacher gives me the right answer.
   B. I have to try to say the right answer.
   C. Both A & B.
   D. I do not like to be corrected

Please explain

_____

_____

Appendix E
Communicative Task 1: Information Gap

Researcher: We are going to start with a small game. Please take this chart, and I will show you how to play. First, do you have any vocabulary questions about the words in the chart or in the word bank below?
Participant: No.
Researcher: O.K., I am going to tell you about one of the items on our charts that have changed human life in some way. I will give you a few hints, and you try to find the item that I am talking about on your chart.So here are some hints. It was invented in the USA. It is used for communication.
Participant: Telephone?
Researcher: Yes, good, now can you tell me about the "when" box because I am missing that information.
Participant: 1876
Researcher: Sorry, it would be more polite to use a sentence.
Participant: Sorry, it was invented in 1876.
Researcher: Oh, could you also tell me about "who"?
Participant: Yes, it was invented by Alexander Graham Bell.
Researcher: O.K. Thanks. Let's try one more. O.K., this one was also created in the US. In fact, it was created in 2004.
Participant: Facebook.
Researcher: Right? Um but could you tell about the last category
Participant: This one? Current use?
Researcher: Yes.
Participant: It is used for communication.
Researcher: O.K. Now your turn, please tell me about a few items and I will guess which one you are talking about. Remember, you can use any of the words in the word bank to help you out.
Participant: O.K. It was built in 1973.
Researcher: Sorry, I need more help.
Participant: It was built in Canada. It was built by the CN railway company.
Researcher: Oh, O.K. the CN Tower right?
Participant: Yes.
Researcher: Sorry, could you tell me about the current use?
Participant: It is used for tourism.
Researcher: Thanks. OK lets try a few more.

WORD BANK
paint, make, create, invent,
build, design, play, use,
discover, write

Researcher

What?	Who?	Where?	When?	Current use?
Paper		China		communication
Pyramids			2500 BC	tourism
Facebook	Mark Zuckerberg	USA	2004	
Telephone		USA		communication
Light bulb	Thomas Edison			illumination
Penicillin	Alexander Flemming		1928	
Basketball		USA	1891	
Superman	Jerry Siegel and Joe Shuster			entertainment
CN Tower	CN Railway company	Canada	1973	
Airplane		USA	1904	
The Simpsons	Matt Groening			entertainment
iPod		USA		entertainment
The Starry Night			1889	pleasure
Hamlet		England	1599	
Mona Lisa	Leonardo da Vinci	Italy		pleasure
Gravity	Newton			balance
Star Wars		USA		entertainment
Happy Birthday song	Patty & Mildred Hill	USA		
Taj Mahal	Shah Jahan			tourism
War & Peace			1869	entertainment

Participant

What	Who	Where	When	Why (now)?
Paper	Ts'ai Lun	China	105	communication
Pyramids	The Egyptians	Egypt	2500 BC	tourism
Facebook	Mark Zuckerberg	USA	2004	communication
Telephone	Alexander Graham Bell	USA	1876	communication
Light bulb	Thomas Edison	USA	1879	illumination
Penicillin	Alexander Flemming	Scotland	1928	medicine
Basketball	James Naismith	USA	1891	fun
Superman	Jerry Siegel and Joe Shuster	USA	1932	entertainment
CN Tower	CN Railway company	Canada	1973	tourism
Airplane	Orville and Wilbur Wright	USA	1904	transportation
The Simpsons	Matt Groening	USA	1987	entertainment
iPod	Apple Computer	USA	2001	entertainment
The Starry Night	Vincent Van Gogh	France	1889	pleasure
Hamlet	William Shakespeare	England	1599	education
Mona Lisa	Leonardo da Vinci	Italy	1503	pleasure
Gravity	Newton	England	1687	balance
Star Wars	George Lucas	USA	1977	entertainment
Happy Birthday song	Patty & Mildred Hill	USA	1893	celebration
Taj Mahal	Shah Jahan	India	1632	tourism
War & Peace	Leo Tolstoy	Russia	1869	entertainment

Appendix F
Communicative Task 2: Story Retelling Task (sample excerpt)

I will play a recording of a story while you look at the word bank and story pictures.

Please **DO NOT** write any notes.

Please save any questions until **AFTER** you hear the story.

After you have heard the story (and asked any questions), please tell me the story.
You may find the pictures and some of the words in the word bank helpful.

WORD BANK

scramble, at home, eat, customers, display, supermarket, wash, bring, then, collect, load, after that, sort, next, package, egg processing plant


Egg Processing Plant

Appendix G
Communicative Task 3: Role Play

You are a real estate agent trying to sell this house at 100 Maple Street.

100 Maple Street: Front Yard in 2010

100 Maple Street: Back Yard in 2010

I am looking for a house. I saw 100 Maple Street in 2010, and it was very bad. I told my wife and daughter that I would never buy that house.

BUT many changes have been made to 100 Maple Street since 2010. Now, in 2012, 100 Maple Street is a very nice house as you can see from the pictures and the chart of changes below. You believe I should buy the house.

100 Maple Street: Front Yard Today

100 Maple Street: Back Yard Today

Changes that have improved 100 Maple Drive

new driveway	no more old car on the lawn	new TV	garden swing	broken pipes
no more window problem	privacy fence	a flower garden	new washer and dryer	no more cockroaches
clean walls, new paint	new swimming pool	a mail box	new reduced price	new bathtub
a lot of plants	no more chimney problem	new curtains	no more rats	

100 Maple Street in 2010 **100 Maple Street Today**

....

You should start our conversation with, "Mr. Quinn, you should buy 100 Maple Street."

You may find the words in the word bank helpful in selling me the house.

WORD BANK

add, repair, paint, remove, build, fix, install, buy, make, exterminate, clean, replace

Appendix H
Error Mark Off Sheet

Task 1 Info Gap

Number	Item	WH Q			
1		WHO,	WHERE,	WHEN,	WHY
2		WHO,	WHERE,	WHEN,	WHY
3		WHO,	WHERE,	WHEN,	WHY
4		WHO,	WHERE,	WHEN,	WHY

Task 2 Story Retell

Number	Item
1	
2	
3	
4	

Task 3 Role Play

Number	Item	Verb
1		
2		
3		
4		

Appendix I
Uniform Immediate and Delayed CF by Task

Immediate and Delayed CF for the Same Error in the Information Gap Task

Immediate CF.

Participant: *Paper invented in 105*

Researcher: [Researcher records error on the error mark off sheet.] *Can you try to tell me about this one again?* [Researcher points to the corresponding gap in the researcher's table.]

Participant: *Paper invent in 105*

Researcher: *Okay, we say, "Paper was invented in 105." Could you repeat that please?*

Participant: *Paper was invented in 105.*

Researcher: [Researcher continues with the information gap task.]

Delayed CF.

Participant: *Paper invented in 105*

Researcher: [Researcher records error on the error mark off sheet continues with the information gap task. After the information gap task is complete the researcher begins CF provision.] *Can you try to tell me about this one again?* [Researcher points to the corresponding gap in the researcher's table.]

Participant: *Paper invent in 105*

Researcher: *Okay, we say, "Paper was invented in 105." Could you repeat that please?*

Participant: *Paper was invented in 105.*

Researcher: [Researcher begins the next provision of CF.]

Immediate and Delayed CF for the Same Error in the Story Retelling Task.

Immediate CF.

Participant: *Then, the eggs are clean.*

Researcher: [Researcher records error on the error mark off sheet.] *Can you try to tell me about this one again?* [Researcher points to the picture of the eggs being cleaned.]

Participant: *Then, the eggs are cleaning*

Researcher: *Okay, we say, 'The eggs are cleaned." Could you repeat that please?*

Participant: *The eggs are cleaned.*

Researcher: [Researcher allows the participant to continue retelling the story.]

Delayed CF.

Participant: *Then, the eggs are clean.*

Researcher: [Researcher records error on the error mark off sheet without interrupting the participant's retelling of the story. After the participant has completed retelling the story, the researcher begins CF provision.] *Can you try to tell me about this one again?* [Researcher points to the points to the picture of the eggs being cleaned.]

Participant: *The eggs are cleaning*

Researcher: *Okay, we say, 'The eggs are cleaned." Could you repeat that please?*

Participant: *The eggs are cleaned.*

Researcher: [Researcher begins the next provision of CF.]

Immediate and Delayed CF for the Same Error in the Role Play Task.

Immediate CF.

Participant: *Wait! House change a lot, like, the windows have repaired.*

Researcher: [Researcher records error on the error mark off sheet.] *Can you try to tell me about this one again?* [Researcher points to the windows in the picture of the repaired house.]

Participant: *The windows have repairs.*

Researcher: *Okay, we say, 'The windows have been repaired." Could you repeat that please?*

Participant: *The windows have been repaired*

Researcher: [Researcher continues with the information gap task.]

Delayed CF.

Participant: *Wait! House change a lot, like, the windows have repaired.*

Researcher: [Researcher records error on the error mark off sheet continues with the role play task. After the role play task is complete the researcher begins CF provision.] *Can you try to tell me about this*

one again? [Researcher points to the windows in the picture of the repaired house.]

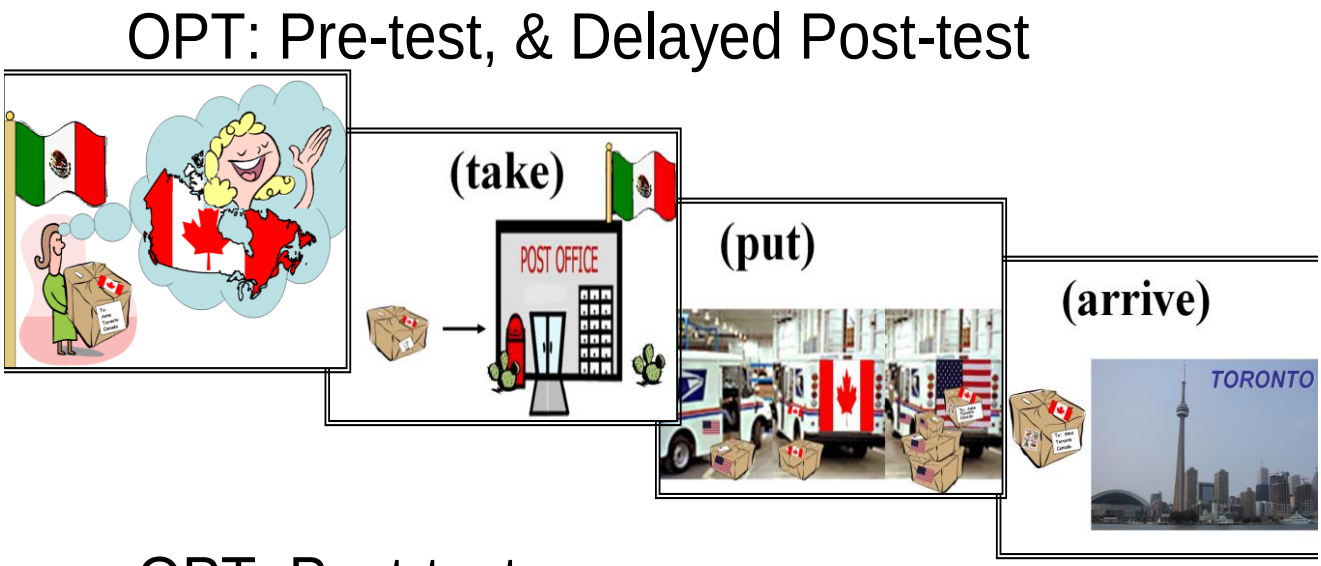
Participant: *The windows have repairs.*

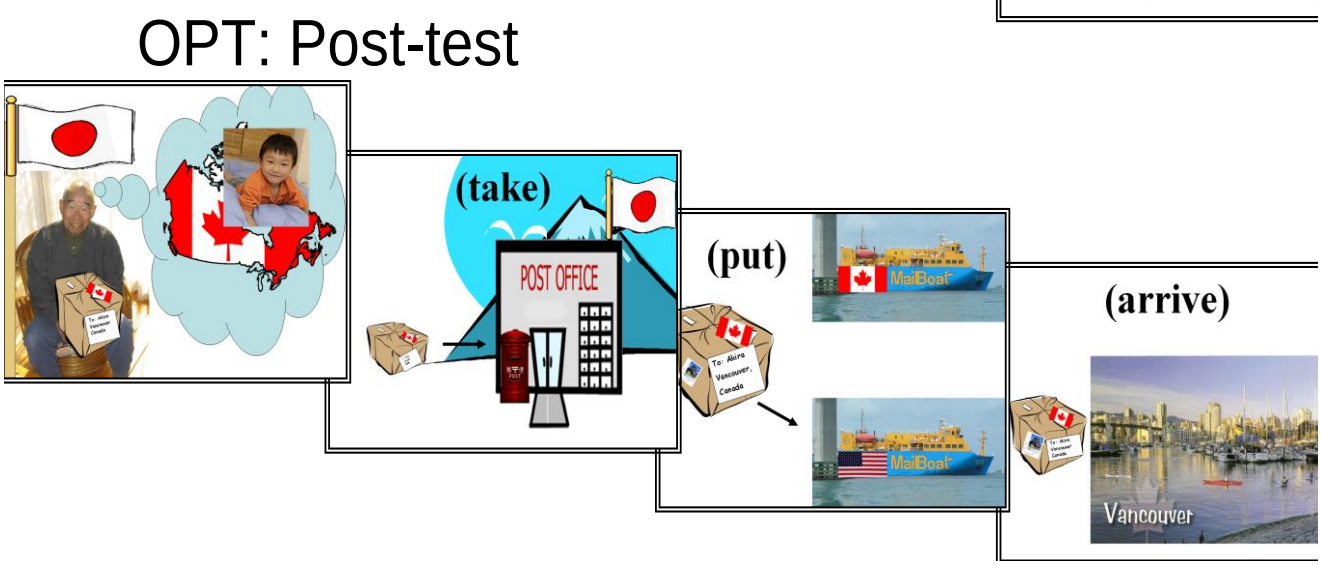
Researcher: *Okay, we say, 'The windows have been repaired." Could you repeat that please?*

Participant: *The windows have been repaired.*

Researcher: [Researcher begins the next provision of CF.]

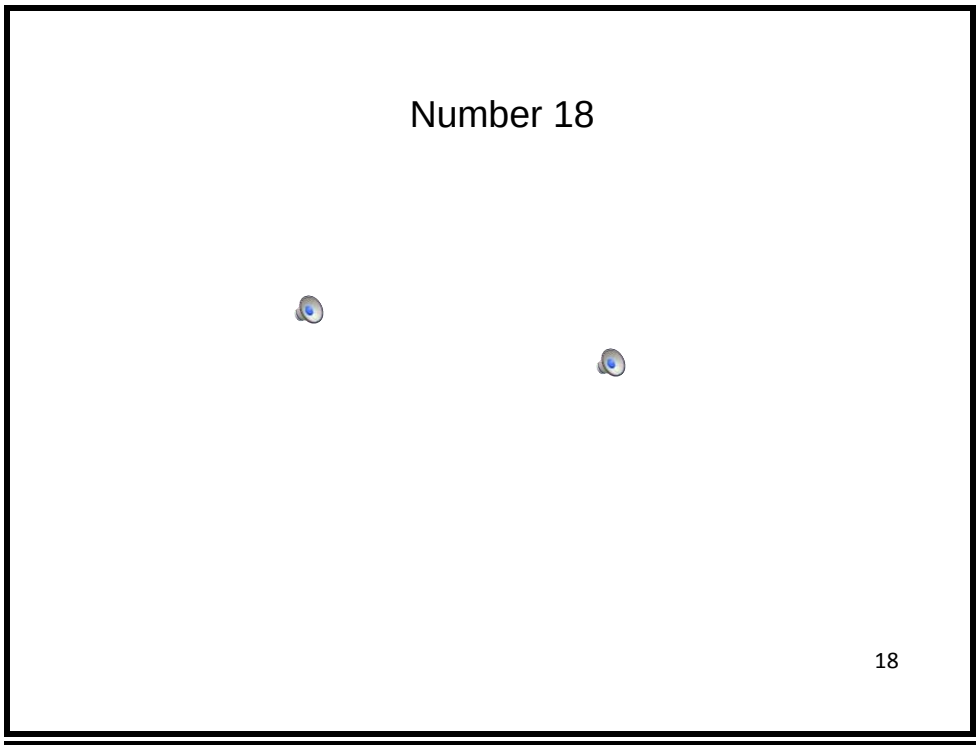
Appendix J
Oral Production Test (sample excerpt)





Appendix K
Aural Grammaticality Judgment Test (sample excerpt)

Participants hear <u>but do not see</u> each item on a PowerPoint slide. After 3 seconds, PowerPoint automatically advances to the next slide.



18. "*The big game was win in overtime."*

Participants must mark correct, incorrect, or not sure on their papers.

<u>Instructions</u>:

Please put a check (√) beside your answer.

1. Correct ___√_ Incorrect _____ Not Sure _____

2. Correct _____ Incorrect ___√___ Not Sure _____

3. Correct _____ Incorrect ___√___ Not Sure _____

...

18. Correct _____ Incorrect ___√___ Not Sure _____

Appendix L
Written Error Correction Test (sample excerpt)

Pre Error Correction Test

Instructions:
This error correction test consists of 24 items. On the following pages, each of the items contains ONLY ONE mistake. Identify the mistake, correct it and explain your correction.

For example,

1. Everybody know that teenagers like to play computer games.

   The ungrammatical part is *know*

   The correct form should be *knows*

   It is ungrammatical because *"Everybody" is a singular noun subject, which needs to be followed with a singular verb form"*

2. The book give to John yesterday.

   The ungrammatical part is *give*

   The correct form should be *was given*

   It is ungrammatical because *"The subject noun phrase "The book" is a receiver of an action (give), which requires a passive verb form following it"*

Appendix M
AGJT and ECT Item Distribution

Aural Grammaticality Judgment Test Item Distribution				
Non-Passive Distractor Items				
Accurate	3			
Incorrect	3			
Passive Target Items		simple present	simple past	present perfect
Accurate	regular	3	3	3
	irregular	3	3	3
Incorrect				
Error of using "ing" (e.g., Apples are picking in the fall.)	regular	1	1	1
	irregular	1	1	1
Error of be-verb omission (e.g., Many bridges built in the 1950's.)	regular	1	1	1
	irregular	1	1	1
Error of improper past participle formation (e.g., Pets have been keep for hundreds of years.)	regular	1	1	1
	irregular	1	1	1

Error Correction Test Item Distribution				
Non-Passive Distractor Items	6			
Passive Target Items		simple present	simple past	present perfect
Error of using "ing" (e.g., Apples are picking in the fall.)	regular	1	1	1
	irregular	1	1	1
Error of be-verb omission (e.g., Many bridges built in the 1950's.)	regular	1	1	1
	irregular	1	1	1
Error of improper past participle formation (e.g., Pets have been keep for hundreds of years.)	regular	1	1	1
	irregular	1	1	1

Appendix N:

Supplementary Non-parametric Analyses

This Appendix describes the non-parametric tests that support the same conclusions of the parametric analyses: All conditions increased their scores statistically significantly from the pre- to post-test and from the pre- to delayed post-tests, and no statistically significant differences were found between the conditions at pre-, post-, or delayed post-tests. The analyses for each language measure (i.e., OPT, AGJT, and ECT) are reported separately. For each language measure, the results for the within group comparisons over time are reported for each CF treatment condition (i.e., No CF, Delayed CF, and Immediate CF). Then, the between groups comparisons at each testing interval (i.e., pre-, immediate post-, and delayed post-test) are reported.

OPT

Within Group Analyses Over Time

No CF Condition

A Friedman test was conducted to evaluate the differences in scores within the No CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 14.47$, $p < .001$. The Kendall coefficient of concordance of .24 indicated moderately strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p = .002$. The mean for the delayed post-test scores was statistically significantly greater than the mean for pre-test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = 2.10$.

Delayed CF Condition

A Friedman test was conducted to evaluate the differences in scores within the Delayed CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 45.49$, $p < .001$. The Kendall coefficient of concordance of .76 indicated strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post-test scores was also statistically significantly greater than the mean for pre-test scores $p < .001$. The mean for the immediate delayed post-test scores was also statistically significantly greater than the mean for delayed post-test scores $p = .015$. (This is the first of three findings that were not revealed in the parametric analyses. The three new findings do not challenge the main conclusions of the parametric analyses: The scores for all conditions increased statistically significantly over time, but no condition

improved statistically significantly more than any other conditions. As was discussed in Chapter 4, the high score for the Delayed CF condition at the immediate post-test appears to be the cause of the group and time interaction found by the mixed-design ANOVA, but not by the follow-up one-way ANOVA.)

Immediate CF Condition

A Friedman test was conducted to evaluate the differences in scores within the Immediate CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 27.06$, $p < .001$. The Kendall coefficient of concordance of .45 indicated moderately strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post-test scores was also statistically significantly greater than the mean for pre-test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = .53$.

Between Groups Analyses at Pre-, Immediate Post, and Delayed Post-Testing

Pre-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the pre-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 4.72$, $p = .09$. Therefore, there were no differences between the conditions at the pre-test.

Immediate Post-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the immediate post-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 4.91$, $p = .09$. Therefore, there were no differences between the conditions at the immediate post-test.

Delayed Post-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the delayed post-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 0.84$, $p = .66$. Therefore, there were no differences between the conditions at the delayed post-test.

AGJT

Within Group Analyses Over Time

No CF Condition

A Friedman test was conducted to evaluate the differences in scores within the No CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 21.30$, $p < .001$. The Kendall coefficient of concordance of .36 indicated moderately strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p = .004$. The mean for the delayed post-test scores was statistically significantly greater than the mean for pre-test scores $p < .001$. The mean for the delayed post-test scores was also statistically significantly greater than the mean for immediate post-test scores $p = .006$.

Delayed CF Condition

A Friedman test was conducted to evaluate the differences in scores within the Delayed CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 21.84$, $p < .001$. The Kendall coefficient of concordance of .36 indicated moderately strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post test scores was also statistically significantly greater than the mean for pre test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = .62$. (This is the second of the three findings that were not revealed in the parametric analyses. In the parametric analysis, all conditions appeared to increase their scores statistically significantly from the immediate post-test to the delayed post-test on the AGJT. As noted above, this new finding does not challenge the main conclusions of the parametric analyses.)

Immediate CF Condition

A Friedman test was conducted to evaluate the differences in scores within the Immediate CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 21.69$, $p < .001$. The Kendall coefficient of concordance of .36 indicated moderately strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post test scores was also statistically significantly greater than the mean

for pre-test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = .55$. (This is the third of the three findings that were not revealed in the parametric analyses. As noted above, in the parametric analyses, all conditions appeared to increase their scores statistically significantly from the immediate post-test to the delayed post-test on the AGJT, but again, this new finding does not challenge the main conclusions of the parametric analyses.)

Between Groups Analyses at Pre-, Immediate Post, and Delayed Post-Testing

Pre-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the pre-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 1.26$, $p = .53$. Therefore, there were no differences between the conditions at the pre-test.

Immediate Post-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the immediate post-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 0.73$, $p = .69$. Therefore, there were no differences between the conditions at the immediate post-test.

Delayed Post-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the delayed post-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 0.60$, $p = .74$. Therefore, there were no differences between the conditions at the delayed post-test.

ECT

Within Group Analyses Over Time

No CF Condition

A Friedman test was conducted to evaluate the differences in scores within the No CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 27.10$, $p < .001$. The Kendall coefficient of concordance of .45 indicated moderately strong differences in scores

among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post test scores was statistically significantly greater than the mean for pre-test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = .93$.

Delayed CF Condition

A Friedman test was conducted to evaluate the differences in scores within the Delayed CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 23.45$, $p < .001$. The Kendall coefficient of concordance of .39 indicated moderately strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post-test scores was also statistically significantly greater than the mean for pre-test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = .22$.

Immediate CF Condition

A Friedman test was conducted to evaluate the differences in scores within the Immediate CF condition over time (i.e., pre-test, immediate post-test, and delayed post-test). The Friedman test was statistically significant, $X^2(2, N=30) = 35.81$, $p < .001$. The Kendall coefficient of concordance of .60 indicated strong differences in scores among the pre-, immediate post-, and delayed post-tests. Follow up analysis with Wilcoxon sign-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .017$. The mean for the immediate post-test scores was statistically significantly greater than the mean for the pre-test scores, $p < .001$. The mean for the delayed post-test scores was also statistically significantly greater than the mean for pre-test scores $p < .001$. However, there was no statistically significant difference between the mean scores for the immediate post-test and delayed post-tests, $p = .03$.

Between Groups Analyses at Pre-, Immediate Post, and Delayed Post-Testing

Pre-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the pre-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 2.43$, $p = .30$. Therefore, there were no differences between the conditions at the pre-test.

Immediate Post-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the immediate post-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 5.35$, $p = .07$. Therefore, there were no differences between the conditions at the immediate post-test.

Delayed Post-test

A Kruskal-Wallis test was conducted to determine whether there was a difference between CF treatment conditions (i.e., No CF, Delayed CF, and Immediate CF) at the delayed post-test. The significance level was set at $p < .05$.The test was not statistically significant $X^2(2, N=90) = 0.48$, $p = .79$. Therefore, there were no differences between the conditions at the delayed post-test.